

JOURNEY SCRAPBOOK

Foundational Training – week 1

14 August, 2023 – 18 August, 2023

Sanat Mahajan

Batch 6

DAY 1

Introduction to BA

Requirement Analysis

Project Management

Waterfall vs Agile Method

Sprint

Scrums and different roles in it

LEARNINGS

In this day I learned about importance of business analyst in an organization also got to know about the SDLC (Software Development Life Cycle) and steps involved in it. Also I learned about Requirement Analysis and different types of requirements like functional and non functional. After this we were explained about waterfall and agile methodologies about their benefits and disadvantages. Later we were told about sprints and scrum and we did a small activity in which we were divided into teams and we were suppose to have sprint planning, scrum meeting, review and retrospective which helped in understanding the working of the methodology. And at last we were showed to use Azure DevOps how to make tasks, epics and user story in it and were explained about these too. With this the session concluded and overall it was a very informative and interactive session. Had a good time in the session.

JOURNEY SCRAPBOOK

Foundational Training – Week 2

21 August, 2023 – 28 August, 2023

Sanat Mahajan

Batch 6

DAY 2

Creating Epic, user story & tasks

Story point estimation

Sprint board

Product / Sprint backlog

Workflow

Burndown, burnup and velocity charts

Screenshots

A screenshot of a web browser displaying the Azure DevOps Boards page for the 'Group Task' project. The URL in the address bar is https://dev.azure.com/mahajansanat15/Group%20Task/_boards/board/t/Group%20Task%20Team/Stories. The page title is 'Group Task'. The left sidebar shows navigation options like Overview, Boards, Work items, Backlogs, Sprints, Queries, Delivery Plans, Analytics views, Repos, Pipelines, Test Plans, and Project settings. The 'Boards' option is selected. The main content area shows a board titled 'Group Task Team' with sections for New, Active, Resolved, and Closed items. Under 'New', there are three items: '11 Recent items' (Unassigned, State: New, 3), '12 Categories' (Unassigned, State: New, 8), and '13 clearance sale' (Unassigned, State: New, 5). A search bar at the top right contains the text 'Search'.

A screenshot of the Azure DevOps Boards Hub. The URL in the address bar is https://dev.azure.com/mahajansanat15/Group%20Task/_boards. The page title is 'Boards'. The left sidebar shows navigation options like Overview, Boards, Work items, Backlogs, Sprints, Queries, Delivery Plans, Analytics views, Repos, Pipelines, Test Plans, and Project settings. The 'Boards' option is selected. The main content area shows a message: 'Try the New Boards Hub for improved performance, accessibility, and new features. [Click here](#) to learn more.' Below this, it says 'Group Task Team' with a star icon and a gear icon. It shows tabs for 'Board' (selected), 'Analytics', and 'View as Backlog'. On the right, there are filters for 'Stories' (dropdown), 'Analytics', 'Queries', and 'Sprints'. Below these are sections for 'New' (11 Recent items, Unassigned, State: New, 3), 'Active' (0/5), 'Resolved' (0/5), and 'Closed' (0/5). A large 'Upcoming' section is visible on the right side.

Azure DevOps mahajansanat15 / Group Task / Boards / Backlogs

Search

GT Group Task + Try the New Boards Hub for improved performance, accessibility, and new features. [Click here to learn more.](#)

Overview

Boards

Work items

Boards

Backlogs

Sprints

Queries

Delivery Plans

Analytics views

Repos

Pipelines

Test Plans

Project settings

Group Task Team star g8

Backlog Analytics + New Work Item View as Board Column Options ... Stories grid filter gear refresh

Order	Work Item Type	Title
1	User Story	> clearance sale
2	User Story	> Recent items
3	User Story	> Categories

Planning
Drag and drop work items to include them in a sprint.

Group Task Team Backlog

Iteration 1 Current
Planned Effort: 5
1 5

Iteration 2
Planned Effort: 8
1 2

Iteration 3
Planned Effort: 3
1 2

https://dev.azure.com/mahajansanat15/Group Task/_backlogs/directory

DAY 3

DBMS

SQL

Normalization

Joins

Select

SubQuery

SCREENSHOTS

```
root@ip-10-0-24-254: /home/ubuntu
File Edit View Search Terminal Help
training=# \l
          List of databases
  Name   |  Owner   | Encoding | Collate |  Ctype  | Access privileges
-----+-----+-----+-----+-----+-----+
postgres | postgres | UTF8    | C.UTF-8 | C.UTF-8 |
template0 | postgres | UTF8    | C.UTF-8 | C.UTF-8 | =c/postgres      +
           |          |          |          |          | postgres=CTc/postgres
template1 | postgres | UTF8    | C.UTF-8 | C.UTF-8 | =c/postgres      +
           |          |          |          |          | postgres=CTc/postgres
training  | postgres | UTF8    | C.UTF-8 | C.UTF-8 |
(4 rows)

training=# create table guest (gid int, mob bigint, email varchar(50), aadhaar bigint);
CREATE TABLE
training=# insert into guest values (1, 1234567890, 'abx@gmail.com', 23456123);
INSERT 0 1
training=# insert into guest values (2, 9876554322, 'abc@gmail.com', 23456345);
INSERT 0 1
training=# create table room (gid int, room_id int, days int);
CREATE TABLE
training=# insert into room values (1, 10, 4);
INSERT 0 1
training=# insert into room values (2, 15, 2);
INSERT 0 1
training=# select * from guest
training-# ;
  gid |  mob   |  email   |  aadhaar
-----+-----+-----+-----+
  1 | 1234567890 | abx@gmail.com | 23456123
  2 | 9876554322 | abc@gmail.com | 23456345
(2 rows)

training=#
```

3.23.95.174 - Remote Desktop Connection

```
+-----+-----+-----+-----+-----+
| postgres | postgres | UTF8   | C.UTF-8 | C.UTF-8 |
| template0 | postgres | UTF8   | C.UTF-8 | C.UTF-8 | =c/postgres      +
|           |           |         |         |         | postgres=CTc/postgres
| template1 | postgres | UTF8   | C.UTF-8 | C.UTF-8 | =c/postgres      +
|           |           |         |         |         | postgres=CTc/postgres
| training  | postgres | UTF8   | C.UTF-8 | C.UTF-8 |
+-----+-----+-----+-----+-----+
(4 rows)

training=# create table guest (gid int, mob bigint, email varchar(50), aadhaar bigint);
CREATE TABLE
training=# insert into guest values (1, 1234567890, 'abx@gmail.com', 23456123);
INSERT 0 1
training=# insert into guest values (2, 9876554322, 'abc@gmail.com', 23456345);
INSERT 0 1
training=# create table room (gid int, room_id int, days int);
CREATE TABLE
training=# insert into room values (1, 10, 4);
INSERT 0 1
training=# insert into room values (2, 15, 2);
INSERT 0 1
training=# select * from guest
training-# ;
      gid |    mob    |      email     | aadhaar
-----+-----+-----+-----+
  1 | 1234567890 | abx@gmail.com | 23456123
  2 | 9876554322 | abc@gmail.com | 23456345
(2 rows)

training=# select * from guest cross join room;
      gid |    mob    |      email     | aadhaar | gid | room_id | days
-----+-----+-----+-----+-----+-----+-----+
  1 | 1234567890 | abx@gmail.com | 23456123 |  1 |      10 |    4
  2 | 9876554322 | abc@gmail.com | 23456345 |  1 |      10 |    4
  1 | 1234567890 | abx@gmail.com | 23456123 |  2 |      15 |    2
  2 | 9876554322 | abc@gmail.com | 23456345 |  2 |      15 |    2
(4 rows)

training=# select * from guest inner join room on guest.gid = room.gid;
      gid |    mob    |      email     | aadhaar | gid | room_id | days
-----+-----+-----+-----+-----+-----+-----+
  1 | 1234567890 | abx@gmail.com | 23456123 |  1 |      10 |    4
  2 | 9876554322 | abc@gmail.com | 23456345 |  2 |      15 |    2
(2 rows)

training=#

```

3.23.95.174 - Remote Desktop Connection

```
training=# create table room (gid int, room_id int, days int);
CREATE TABLE
training=# insert into room values (1, 10, 4);
INSERT 0 1
training=# insert into room values (2, 15, 2);
INSERT 0 1
training=# select * from guest
training-# ;
+-----+-----+-----+
| gid | mob | email | aadhaar |
+-----+-----+-----+
| 1 | 1234567890 | abx@gmail.com | 23456123 |
| 2 | 9876554322 | abc@gmail.com | 23456345 |
+-----+
(2 rows)

training=# select * from guest cross join room;
+-----+-----+-----+-----+-----+-----+
| gid | mob | email | aadhaar | gid | room_id | days |
+-----+-----+-----+-----+-----+-----+
| 1 | 1234567890 | abx@gmail.com | 23456123 | 1 | 10 | 4 |
| 2 | 9876554322 | abc@gmail.com | 23456345 | 1 | 10 | 4 |
| 1 | 1234567890 | abx@gmail.com | 23456123 | 2 | 15 | 2 |
| 2 | 9876554322 | abc@gmail.com | 23456345 | 2 | 15 | 2 |
+-----+
(4 rows)

training=# select * from guest inner join room on guest.gid = room.gid;
+-----+-----+-----+-----+-----+-----+
| gid | mob | email | aadhaar | gid | room_id | days |
+-----+-----+-----+-----+-----+-----+
| 1 | 1234567890 | abx@gmail.com | 23456123 | 1 | 10 | 4 |
| 2 | 9876554322 | abc@gmail.com | 23456345 | 2 | 15 | 2 |
+-----+
(2 rows)

training=# insert into room values (1, 14, 7);
INSERT 0 1
training=# insert into room values (1, 4, 1);
INSERT 0 1
training=# insert into room values (2, 18, 1);
INSERT 0 1
training=# select count(gid), gid, room_id, days from room group by gid;
ERROR: column "room.room_id" must appear in the GROUP BY clause or be used in an aggregate function
LINE 1: select count(gid), gid, room_id, days from room group by gid...
^

training=# select count(room_id), gid from room group by gid;
+-----+
| count | gid |
+-----+
| 2 | 2 |
| 3 | 1 |
+-----+
(2 rows)
```

DAY 4

Testing

STLC

Testing Triangle

Automation Testing

Unit Testing

White / Black / Grey Box Testing

SOFTWARE TESTING

- Understood how important Testing is for any product and the difference between bugs, defects and errors.
- Learnt Software Development life cycle and Software testing life cycle with their stages that run in a loop.
- Obtained insights on Unit testing and integration testing and the need to perform both.
- Acquired knowledge on following classes: @BeforeAll, @BeforeEach, @AfterEach, @Order
- Observed Softwares (For writing test cases) : Junit, Fitnesse .

STLC

Test plan – Purpose, Goals, Types, Cycles, Strategy, Schedules, Environment, Criteria. Metrics, Defect types, Risks, etc; will be planned in this stage

Test Requirements – All the requirements and sub requirements of the product are documented

Test case design – Necessary positive and negative test cases for each requirement is generated

-Allocation – Testers and resources are allocated to each requirement based on priority and complexity. Test environment is prepared

Test execution – Actual testing happens

Test evaluation – The validation of testing done

Log defects – Defects need to be recorded

Test Reports – Final report of testing with all the analysis will have to be documented for further correction or deployment

AUTOMATION TESTING

- Tools like Selenium, Cucumber are used for Testing Automation
- Types of development – TDD (Test Driven Dev), BDD (Behavioral Driven Dev)
- BDD Cucumber Framework – provides good communication between teams, works on rapid, small iterations and helps to check system documentation with actual system behavior
- It works on structure of Discovery, Formulation and Automation
- Majorly used for Final Acceptance testing

DAY 5

DevOps

DevOps Life Cycle

DevOps Stages

CI / CD

Github Actions

DEVOPS (DEVELOPMENT + OPERATIONS)

- This domain has the Development team that caters to the planning, designing, coding and testing of a product and the IT operations team verifies the requirements, deploys the product and monitors it.
- These teams go hand in hand and run in a loop.
- Gained an understanding on the seven stages of this domain.
 - continuous development
 - continuous integration
 - continuous testing
 - continuous deployment
 - continuous monitoring
 - continuous feedback
 - continuous operations

GITHUB ACTIONS

18

The screenshot shows a GitHub repository page for 'Unext-training'. The main navigation bar includes links for Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. The 'Actions' tab is selected. On the left, the file tree shows a 'main' branch and a '.github/workflows' folder containing a file named 'github_actions.yml'. The 'github_actions.yml' file is currently selected. The right side of the screen displays the contents of the file:

```
name: github_actions
run-name: ${{ github.actor }} is learning GitHub Actions
on: [push]
jobs:
  check-bats-version:
    runs-on: ubuntu-latest
    steps:
      - uses: actions/checkout@v3
      - uses: actions/setup-node@v3
        with:
          node-version: '14'
      - run: npm install -g bats
      - run: bats -v
```

← X https://github.com/sanatMahajan/Unext-training/actions/runs/5971809100/job... ⌂ A ⭐ ⓘ 🔍 FCM A ⚙ | ⌂ ⚙ ⚙ 🔍

sanatMahajan / Unext-training

Type ⌂ to search

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

← github_actions

✓ sanatMahajan is learning GitHub Actions #1

Re-run all jobs Latest #2 ...

Summary

Jobs

check-bats-version

Run details

Usage

Workflow file

check-bats-version succeeded 5 minutes ago in 4s

Search logs

Set up job 1s

Run actions/checkout@v3 0s

Run actions/setup-node@v3 0s

Run npm install -g bats 0s

Run bats -v 0s

Post Run actions/setup-node@v3 0s

Post Run actions/checkout@v3 0s

Complete job 0s

https://github.com/sanatMahajan/Unext-training/actions/runs/5971809100/job/16201620153

DAY 6

Cloud Computing

Cloud Service Models

Virtualization / Virtual Machines

Cloud Storage

Advantages / Disadvantages of cloud computing

CLOUD COMPUTING

- Cloud Computing is a pretty useful concept to connect, store data, build, deploy and do much more virtually.
- Acquired information on Iaas (Information as a service), Paas (Platform as a service) and Saas (Software as a service).
- Built a Virtual machine using Microsoft Azure cloud services and understood the various configurations that could be used.
- Applied customized disk storage and network settings based on the given requirements.

SCREENSHOTS

22

Microsoft Azure Search resources, services, and docs (G+ /)

Home >

NewVM1

Virtual machine

Search

Connect Start Restart Stop Capture Delete Refresh Open in mobile Feedback CLI / PS

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Resource group ([move](#))
[RDBG](#)

Status
Stopped (deallocated)

Location
East US

Subscription ([move](#))
[RDBG](#)

Subscription ID
23d99a1f-8507-41a7-b40b-64402a51e95e

Health state
-

Tags ([edit](#))
[Add tags](#)

Operating system
Linux

VM availability status
Available

Size
Standard B2s (2 vcpus, 4 GiB memory)

Public IP address
[NewVM1ip337](#)

Virtual network/subnet
[ManiVM-vnet/default](#)

DNS name
[Not configured](#)

Properties Monitoring Capabilities (7) Recommendations Tutorials

Virtual machine

Computer name NewVM1

Operating system Linux (Ubuntu 20.04)

Networking

Public IP address [NewVM1ip337](#) (Network interface [newvm1454](#))

```
[*] login as: azureuser  
[*] Authenticating with public key "imported-openssh-key"  
Welcome to Ubuntu 20.04.6 LTS (GNU/Linux 5.15.0-1042-azure x86_64)
```

```
* Documentation: https://help.ubuntu.com  
* Management: https://landscape.canonical.com  
* Support: https://ubuntu.com/advantage
```

```
System information as of Fri Aug 25 06:46:42 UTC 2023
```

```
System load: 0.01 Processes: 101  
Usage of /: 5.3% of 28.89GB Users logged in: 0  
Memory usage: 30% IPv4 address for eth0: 10.0.0.4  
Swap usage: 0%
```

```
Expanded Security Maintenance for Applications is not enabled.
```

```
0 updates can be applied immediately.
```

```
Enable ESM Apps to receive additional future security updates.  
See https://ubuntu.com/esm or run: sudo pro status
```

```
The list of available updates is more than a week old.  
To check for new updates run: sudo apt update
```

```
The programs included with the Ubuntu system are free software;  
the exact distribution terms for each program are described in the  
individual files in /usr/share/doc/*/*copyright.
```

```
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by  
applicable law.
```

```
To run a command as administrator (user "root"), use "sudo <command>".  
See "man sudo_root" for details.
```

DAY 7

Case Study – Deployment Of Java Spring Boot App

- Azure DevOps
- Azure VM
- Azure Web App Services
- GitHub – Fork, Feature Branch, Pull Request, Merge
- GitHub Actions – Automate Workflow
- Docker – Make Docker Image And Dockerfile
- Maven – Run Maven Package
- Prometheus – To Monitor Local Environment

Name – Sanat Mahajan

Batch - 6

Azure Board Activities:

The screenshot shows the Azure DevOps Boards interface for the 'Case Study 1 Team' board. The left sidebar navigation bar includes links for Overview, Boards, Work items, Boards, Backlogs, Sprints, Queries, Delivery Plans, Analytics views, Repos, Pipelines, and Test Plans. The 'Boards' link is currently selected. The main area displays the backlog under the 'Case Study 1 Team' board. A backlog item titled '3 Deployment of Spring Boot Application' is highlighted with a blue border. The item details show it is 'New', has 'Story Points 3', and is at '11/11' completion. The backlog is organized into columns: New (1 item), Active (0/5), Resolved (0/5), and Closed (0/5). A search bar at the top right contains the placeholder 'Search'.

Welcome back!

Pick up where you left off:

Monday

JOURNEY SCRAPBOOK

Custom Bootcamp – Week 1
29 August, 2023 – 1 September, 2023

Sanat Mahajan

Batch 3

DAY 1

- What Is Data?
- Types Of Data
- What Is Database?
- What Does Data Engineer Do?
- Normalization And Denormalization
- Normalization Forms (1NF,2NF,3NF,BCNF)
- Dependencies (Partial, Transitive, Functional)
- Attribute Keys (Candidate, Primary, Foreign, Super)
- Anomalies- Insert, Update, Delete
- Data Modelling And ER Diagram
- Activity- To Make An ER Diagram For Ordering A Book From Online Store, Perform Normalization
- Dimensional Modelling
- Fact Table And Dimension Table
- Star Schema & Snowflake Schema
- Activity- Implement The Concepts Of Star And Snowflake Schema By Drawing Respective Diagrams
- SCD (Slowly Changing Dimension)



Order

Order ID	Date	Book ID	Customer ID
1	21.08.23	1	30
2	22.08.23	1, 5, 7, 9	31
3	23.08.23	2, 5	32
4	24.08.23	5, 7	33

↓ INF

Order

Order ID	Date	Book ID	Customer ID
1	21.08.23	1	30
2	22.08.23	1	31
2	22.08.23	3	31
2	22.08.23	7	31
2	22.08.23	9	31
3	23.08.23	2	32
3	23.08.23	5	32
4	24.08.23	6	33
4	24.08.23	7	33

Product

Category	cp_id	int
p-type	string	

Date

date	int
------	-----

month

string	
--------	--

year

int	
-----	--

week

int	
-----	--

Product

p_id	int
------	-----

p-name

string	
--------	--

p-type

string	
--------	--

Company

pid	int
-----	-----

p-loc

string	
--------	--

Location

cid	int
-----	-----

area

string	
--------	--

city

string	
--------	--

state

string	
--------	--

Revenue

int	
-----	--

Day

int	
-----	--

Month

string	
--------	--

Year

int	
-----	--

Customer

cid	int
-----	-----

c-name

string	
--------	--

s-name

string	
--------	--

c-add

string	
--------	--

c-phn

int	
-----	--

Revenue

int	
-----	--

DAY 2

- What Is Big Data?
- 4 V's Of Big Data - Volume, Variety, Velocity, Veracity
- Batch And Stream Data Processing.
- Extract, Load, Transform(ETL) Pipelines
- How To Process Big Data?
 - Hadoop
 - Pyspark
- Parallel Processing And Distributive Computing
- Hadoop Used For Batch Processing & Pyspark Used For Stream Processing.
- Types Of Data Structure
 - Structured Data
 - Semi-structured Data
 - Unstructured Data
- Data Warehouse -
 - Save Data As Object
 - Tables
 - Views
 - Structured Data (Fixed Schema-data Allocated Even If Value Is Null)
 - Dimension Modelling
 - ACID Property In DBMS -
 - Atomicity
 - Consistency
 - Isolation
 - Durability
- Datalake -
 - Saves All Types Of Data (Structured, Semi-structured And Unstructured)
 - Flexible Schema (No Data Allocated For Null Values)
 - No ACID Properties

DAY 2

- Data Lakehouse
 - Combination Of Data Warehouse And Datalake.
 - Took Advantages Of Both Data Warehouse And Datalake.
- Creating A VM Using Microsoft Azure.
- What Is Cloud Computing.
- Types Of Cloud -
 - Public Cloud
 - Private Cloud
 - Hybrid Cloud
- Cloud Benefits -
 - High Availability
 - Scalability
 - Agility
 - Fault Tolerance
 - Elasticity
 - Security
 - Disaster Recovery
 - Customer Latency Capabilities
- Compare Capex(capital Expenditure) Vs OpEx(operational Expenditure) -
 - Cloud Is OpEx.
- Infrastructure As A Service(IAAS)
- Platform As A Service(PAAS)
- Software As A Service(SAAS)
- Shared Responsibility Model
- Serverless Computing

DAY 2

- Azure Architecture
 - Regions
 - Region Pairs
 - Availability Zones
- Azure Resources
- Azure Subscriptions -
 - Billing Boundary
 - Access Control Boundary
- Azure Resource Manager
- Resource Groups
- Subscriptions -> Resource Groups

DAY 3

Import favorites | For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure 1 ? ? Shellunext_1693422184...
UNEXT (NPUNEXT.ONMICROSOFT.COM)

Home > Microsoft.SQLDatabase.newDatabaseNewServer_28f552b8ff754d0597909 | Overview ... X

 Microsoft.SQLDatabase.newDatabaseNewServer_28f552b8ff754d0597909 Deployment

« Delete Cancel Redeploy Download Refresh

 Your deployment is complete

 Deployment name : Microsoft.SQLDatabase.ne... Start time : 8/31/2023, 10:28:56 AM
Subscription : npunext-1673505552909 Correlation ID : dd6c9042-f12b-496b-9544...
Resource group : RG_IDAShell_15

> Deployment details
▽ Next steps

[Go to resource](#)

Give feedback 

Tell us about your experience with deployment 

 Cost management
Get notified to stay within your budget and prevent unexpected charges on your bill.
[Set up cost alerts >](#)

 Microsoft Defender for Cloud
Secure your apps and infrastructure
[Go to Microsoft Defender for Cloud >](#)

 Free Microsoft tutorials
Start learning today >

 Work with an expert
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.

 Import favorites | For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

 Microsoft Azure

 Search resources, services, and docs (G+)



Shellunext_1693422184...
UNEXT



Home > sanatdb (sanatdbserver/sanatdb)

sanatdb (sanatdbserver/sanatdb) | Query editor (preview)

SQL database

 Search

«  Login  New Query  Open query  Feedback  Getting started

 Overview

 Activity log

 Tags

 Diagnose and solve problems

 Query editor (preview)

Settings

 Compute + storage

 Connection strings

 Properties

 Locks

Data management

 Replicas

 Sync to other databases

Integrations

sanatdb (sqladmin)

 Query 1

 Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

>  Tables

>  Views

>  Stored Procedures

Query 1

 Run

Cancel query

 Save query

 Export data as

 Show only Editor

1 `select getdate()`

Results

Messages

 Search to filter items...

2023-08-31T06:03:49.8730000

 Query succeeded | 0s

Import favorites | For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure Search resources, services, and docs (G/+) Search Logout

Home > sanatdb (sanatdbserver/sanatdb)

sanatdb (sanatdbserver/sanatdb) | Query editor (preview)

SQL database

Search Login New Query Open query Feedback Getting started

sanatdb (sqladmin)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

- dbo.Author
- dbo.Book
- dbo.customer
- dbo.Orders
- dbo.Publisher

Views

Stored Procedures

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 create table Author (Aid int primary key, A_name varchar(255),no_of_books int)
2 create table Publisher(Pid int primary key,p_name varchar(255),p_country varchar(255))
3 create table Book (bid int primary key,Author_id int foreign key references Author(Aid),Pid int foreign key references Publisher(Pid),B_name varchar(255),B_desc varchar(255),B_qty int)
4 create table Orders (Oid int primary key,odate datetime,bid int foreign key references Book(bid),Cid int foreign key references customer(Cid),O_qty int)
5 create table customer(Cid int primary key,c_name varchar(255),oid int foreign key references Orders(Oid))
```

Results Messages

Query succeeded: Affected rows: 0

Query succeeded | 0s

DAY 4

- Create a database and elastic pool.
- JOINS:
 - Inner Join
 - Left Join
 - Right Join
 - Self Join
 - Outer Join
 - Cross Join
- Stored Procedure
- Views
- Functions
 - Scaler
 - Table Valued
- Complex codes -> Stored Procedures
- Simple code -> Views
- Temporary Table: # before table name creates a temporary table (i.e. no physical table created). It drops automatically.
- Sub-Queries
- UNION and UNION ALL
- Intersection

Import favorites

For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure

Search resources, services, and docs (G+/-)

Shellunext_1693422184...
UNEXT

Home > sanatdb (sanatdbserver/sanatdb)

sanatdb (sanatdbserver/sanatdb) | Query editor (preview)

SQL database

 Login  New Query  Open query  Feedback  Getting started Overview Activity log Tags Diagnose and solve problems Query editor (preview)

Settings

 Compute + storage Connection strings Properties Locks

Data management

 Replicas Sync to other databases

Integrations

Query 1

 Run Cancel query  Save query  Export data as  Show only Editor

```
1  SELECT TOP (1000) * FROM [SalesLT].[Customer]
2  INNER JOIN [SalesLT].[CustomerAddress] ON
3  [SalesLT].[Customer].CustomerID = [SalesLT].[CustomerAddress].CustomerID;
```

Results

 Search to filter items...

CustomerID	NameStyle	Title	FirstName	MiddleName
29485	False	Ms.	Catherine	R.
29486	False	Ms.	Kim	

 Query succeeded | 1s

JOURNEY SCRAPBOOK

Custom Bootcamp – Week 2

4 September, 2023 – 8 September, 2023

Sanat Mahajan

Batch 3

DAY 1

- Create a resource -> Azure storage Account
- Resource Group
- Storage Account
- Storage Services
- Redundancy
 - a. Local redundancy
 - b. Zone redundancy
 - c. Geo redundancy
 - d. Geo zone redundancy
- RAGRS (Read Access Geo Redundant Storage)
 - a. Failover
 - b. Synchronous replication
- Security
- Access protocols

DAY 1

- Access tiers
 - a. Hot
 - b. Cold
 - c. Cool
 - d. Archive Access
- Purging & Archiving
- Containers - Folder for storing files
- Azure Tables
- Azure Queues
- File Shares
- Snapshots
- Ingress & Egress
- Static Website

Import favorites | For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

42

Microsoft Azure ... 3 ? ? Shellunext_1693422184...
UNEXT (NPUNEXT.ONMICROSOFT.COM)

Home > sanatgen2storage_1693809260387 | Overview > sanatgen2storage | Containers >

gen2sancon ...

Container



Upload Add Directory Refresh | Rename Delete Change tier Acquire lease Break lease Give feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Manage ACL

Access policy

Properties

Metadata

Authentication method: Access key ([Switch to Azure AD User Account](#))

Location: gen2sancon

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type
No results				



Presentation title

43

Recycle Bin

File Home Share View Drive Tools Manage sanatfs (\sanatidastorage.file.core.windows.net) (Z:)

← → ⌂ ⌃ ⌄ ⌅ This PC > sanatfs (\sanatidastorage.file.core.windows.net) (Z:) Search sanatfs (\sanatidastorage... 🔎

Name	Date modified	Type	Size
test	9/4/2023 8:59 AM	Text Document	

Quick access

- Desktop
- Downloads
- Documents
- Pictures

This PC

- 3D Objects
- Desktop
- Documents
- Downloads
- Music
- Pictures
- Videos
- Windows (C:)
- Temporary Storage
- sanatfs (\sanatidastorage.file.core.windows.net) (Z:)

Network

1 item



anushkashell | Shared access signature

Storage account

Search



Give feedback

Data migration

Basic networking

Microsoft network routing

Internet routing

Events

Some routing options are disabled because the endpoints are not published.

Storage browser

Signing key

key1

Generate SAS and connection string

Data storage

Containers

Connection string

BlobEndpoint=https://anushkashell.blob.core.windows.net/;QueueEndpoint=https://anushkashell.queue.core.windows.net/

File shares

Queues

Tables

SAS token

?sv=2022-11-02&ss=bfqt&srt=c&sp=rw&lac=upyx&se=2023-09-05T11:50:26Z&st=2023-09-05T03:50:26Z&spr=htt

Security + networking

Networking

Blob service SAS URL

<https://anushkashell.blob.core.windows.net/?sv=2022-11-02&ss=bfqt&srt=c&sp=rw&lac=upyx&se=2023-09-05T11:50:26Z&st=2023-09-05T03:50:26Z&spr=htt>

Access keys

Import favorites | For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Presentation title

Microsoft Azure 1 ? ... Shellunext_1693422184... UNEXT (NPUNEXT.ONMICROSOFT.COM)

Home >

 sankey Key vault

Delete Move Refresh Open in mobile

⚠ Upcoming TLS 1.0, 1.1 deprecation: Please enable support for TLS 1.2 on clients (applications/platform) to avoid any service impact. Learn more here. →

JSON View

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Access policies

Events

Objects

Keys

Secrets

Certificates

Settings

Access configuration

Networking

Essentials

Resource group ([move](#))
[RG_idasanat](#)

Location
East US

Subscription ([move](#))
[npunext-167350552909](#)

Subscription ID
6b56d91d-b19a-47f3-a300-90af21174bd2

Vault URI
<https://sankey.vault.azure.net/>

Sku (Pricing tier)
Standard

Directory ID
dce87315-8ffa-4a01-ab40-0de5a7214b2f

Directory Name
Unext

Soft-delete
[Enabled](#)

Purge protection
[Disabled](#)

Tags ([edit](#))
[Add tags](#)

Get started Properties Monitoring Tools + SDKs Tutorials

Manage keys and secrets used by apps and services

 Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

 Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration.

Set up code repository

Data factory

sanatDF

New ✓



Ingest

Copy data at scale once or
on a schedule.



Orchestrate

Code-free data pipelines.



Configure SSIS

Manage & run your SSIS packages in the cloud.



Home > Microsoft.DataFactory-20230905135340 | Overview >



⋮ ⚡ ⭐ ...

Search

Delete

JSON View

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Networking

Managed identities

Properties

Locks

Getting started

Quick start

Monitoring

Alerts

^ Essentials

Resource group ([move](#))

[RG_idasanat](#)

Status

Succeeded

Location

East US

Subscription ([move](#))

[npunext-1673505552909](#)

Subscription ID

6b56d91d-b19a-47f3-a300-90af21174bd2

Type

Data factory (V2)

Getting started

[Quick start](#)



Azure Data Factory Studio

Launch studio



Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory

Validate all

Publish all 2

Preview experience

Off



Binary1

pipeline1

...



Activities

» <<

✓ Validate

✓ Validate copy runtime

▷ Debug

⚡ Add trigger

{ } ...

Search activities

Move and transform

Copy data

Data flow

Synapse

Notebook

Spark job definition

Azure Data Explorer

Azure Data Explorer C...

Azure Function

Azure Function

Batch Service

Custom

Copy data



Copy data1



General

Source 1

Sink 1

Mapping

Settings

User properties

Name *

Copy data1

Description

Activity state (preview)

 Active Inactive

Timeout

0.00:05:00

DAY 3

- Create data Factory
- Activities in Pipeline :
 - String, Boolean, Array and Integer activities in Pipelines
 - Execute pipeline activity
 - SSIS package (Azure SSIS IR - required for this package)
 - Get Metadata Activity
- Copy Data Activity Implementation
- Parameter Concept Implementation
- Triggers
- Get Metadata Activity
- If-Else Condition for Get Metadata in pipelines.
- For Each loop

Subscription Details | Nuvepro | A output - Microsoft Azure | sanatDF - Azure Data Factory | +

Import favorites For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure | Data Factory > sanatDF metadata X

Shellunext_1693422184838@npunext.onmicrosoft.com UNEXT

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory! X

Data Factory Validate all Publish all 1 Preview experience Off

PL_IDA_01 DS_Source DS_output

Activities Validate Debug Add trigger

set variable

General Set variable

Get Metadata Get Metadata1 Set variable Set variable1

Copy data Copy data1

Parameters Variables Settings Output All status Monitor in Azure Metrics Export to CSV

Showing 1 - 3 of 3 items

Activity name ↑	Activity status ↑↓	Activity type ↑↓	Run start ↑↓	Duration ↑↓	Log
Set variable1	Succeeded	Set variable	9/6/2023, 2:39:51 PM	Less than 1s	
Get Metadata1	Succeeded	Get Metadata	9/6/2023, 2:39:47 PM	3s	
Copy data1	Succeeded	Copy data	9/6/2023, 2:39:37 PM	10s	

2:45 PM 9/6/2023

Subscription Details | Nuvepro X | A output - Microsoft Azure X | sanatDF - Azure Data Factory X | Lumen X | + - □ ×

Import favorites For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure | Data Factory > sanatDF metadata X 1 7 ? 🔍 Shellunext_1693422184838@npunext.onmicrosoft.com UNEXT

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory! X

Data Factory Validate all Publish all Preview experience Off

PL_IDA_01 DS_Source DS_output PL_IDA_02 { } ⚙️ ...

CSV DelimitedText DS_Source

Connection Schema Parameters

Linked service * LS_adlsGen2 Test connection Edit + New Learn more

File path * input / Directory / @dataset().DS_inputpar Browse Preview data Detect

Compression type Select...

Column delimiter ⓘ Comma (,)

Row delimiter ⓘ Default (\r,\n, or \r\n)

Encoding ⓘ Default(UTF-8)



Subscription Details | Nuvepro X | A output - Microsoft Azure X | sanatDF - Azure Data Factory X | create a pipeline in azure devop: X | - | X

Import favorites For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure | Data Factory > sanatDF metadata X

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory Validate all Publish all 1 Preview experience Off

PL_IDA_01 DS_Source DS_output PL_IDA_02

Activities Validate Debug Add trigger

if con

Iteration & conditionals If Condition

Get Metadata Get Metadata1

If Condition If Condition1

True Copy data1

False

Parameters Variables Settings Output

Showing 1 - 3 of 3 items

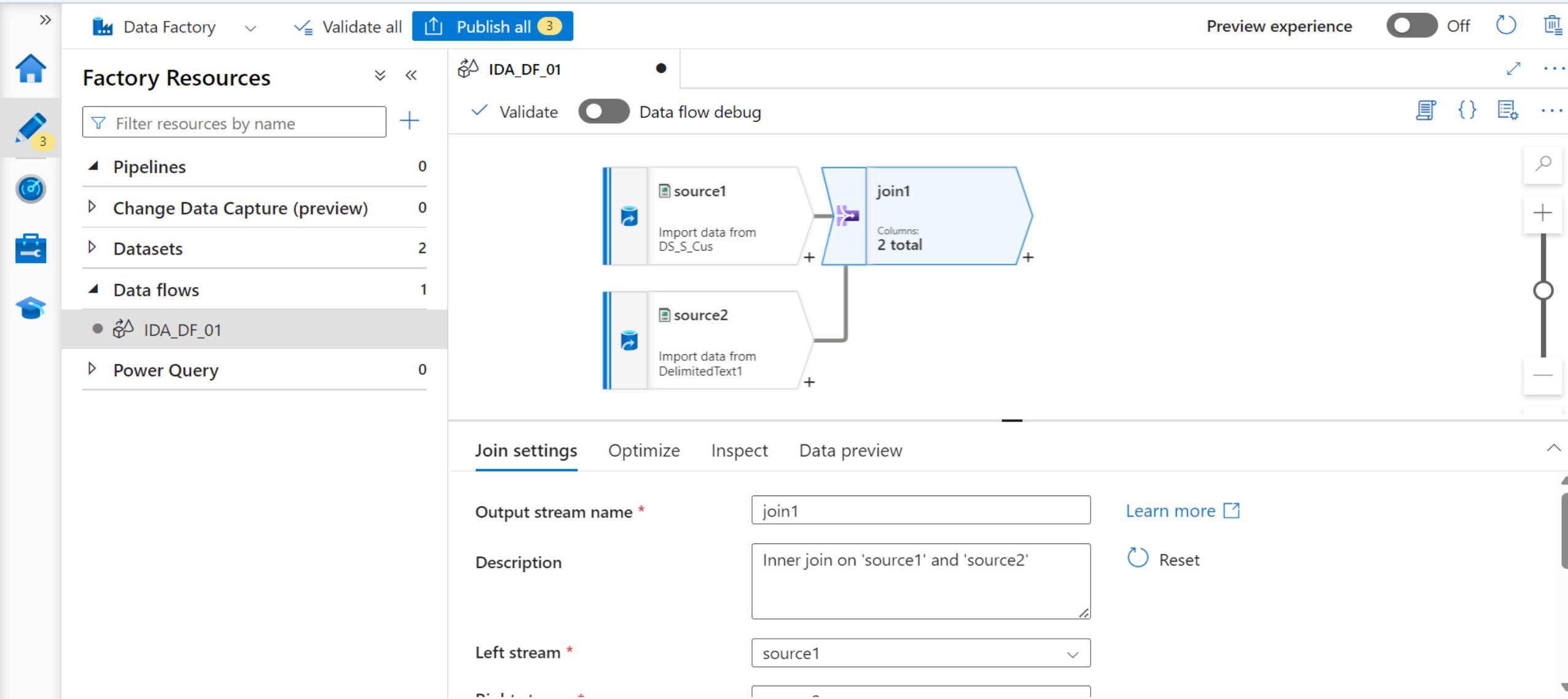
Activity name	Activity status	Activity type	Run start	Duration	Log
Copy data1	Succeeded	Copy data	9/6/2023, 3:01:43 PM	9s	

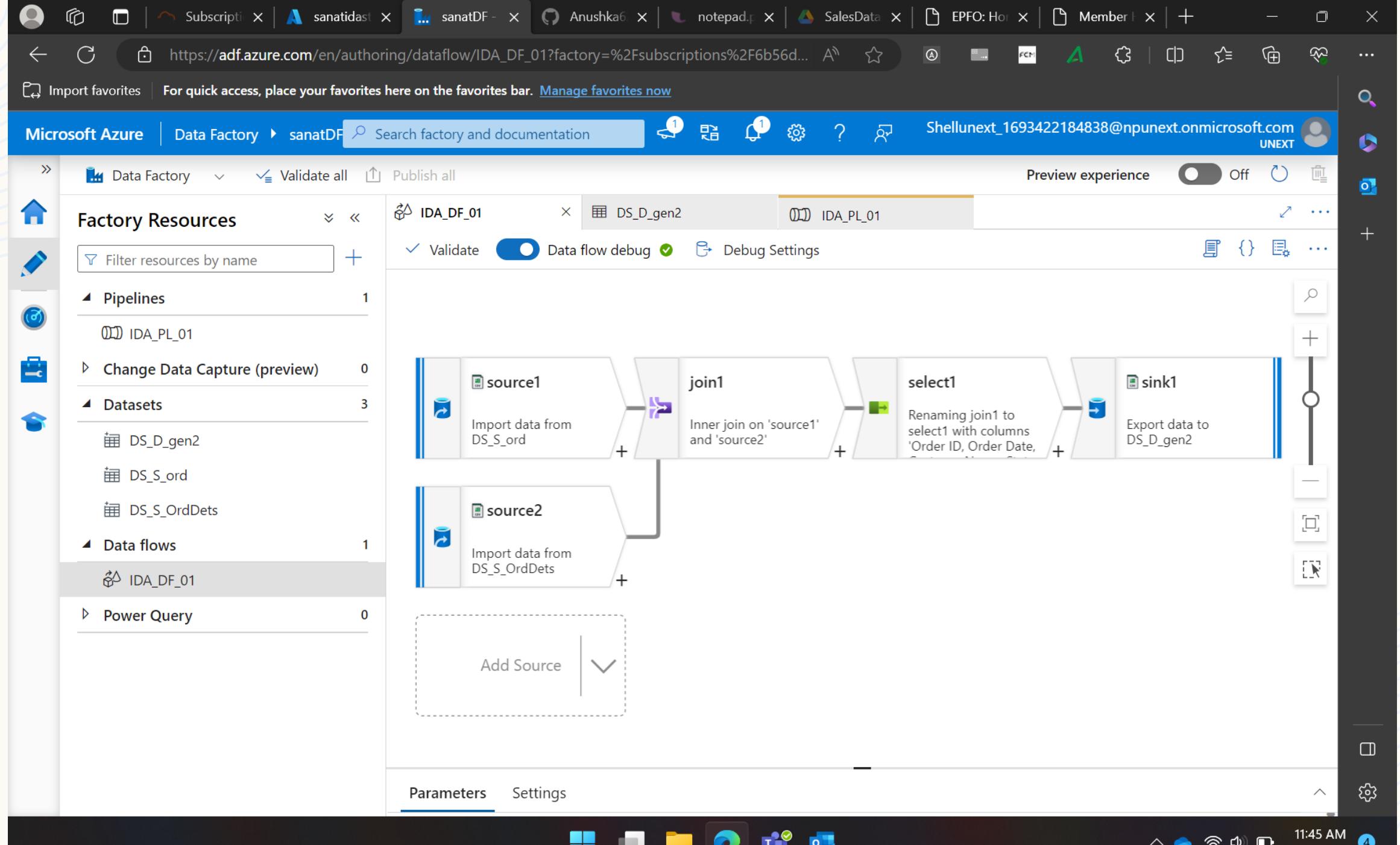
3:03 PM 9/6/2023

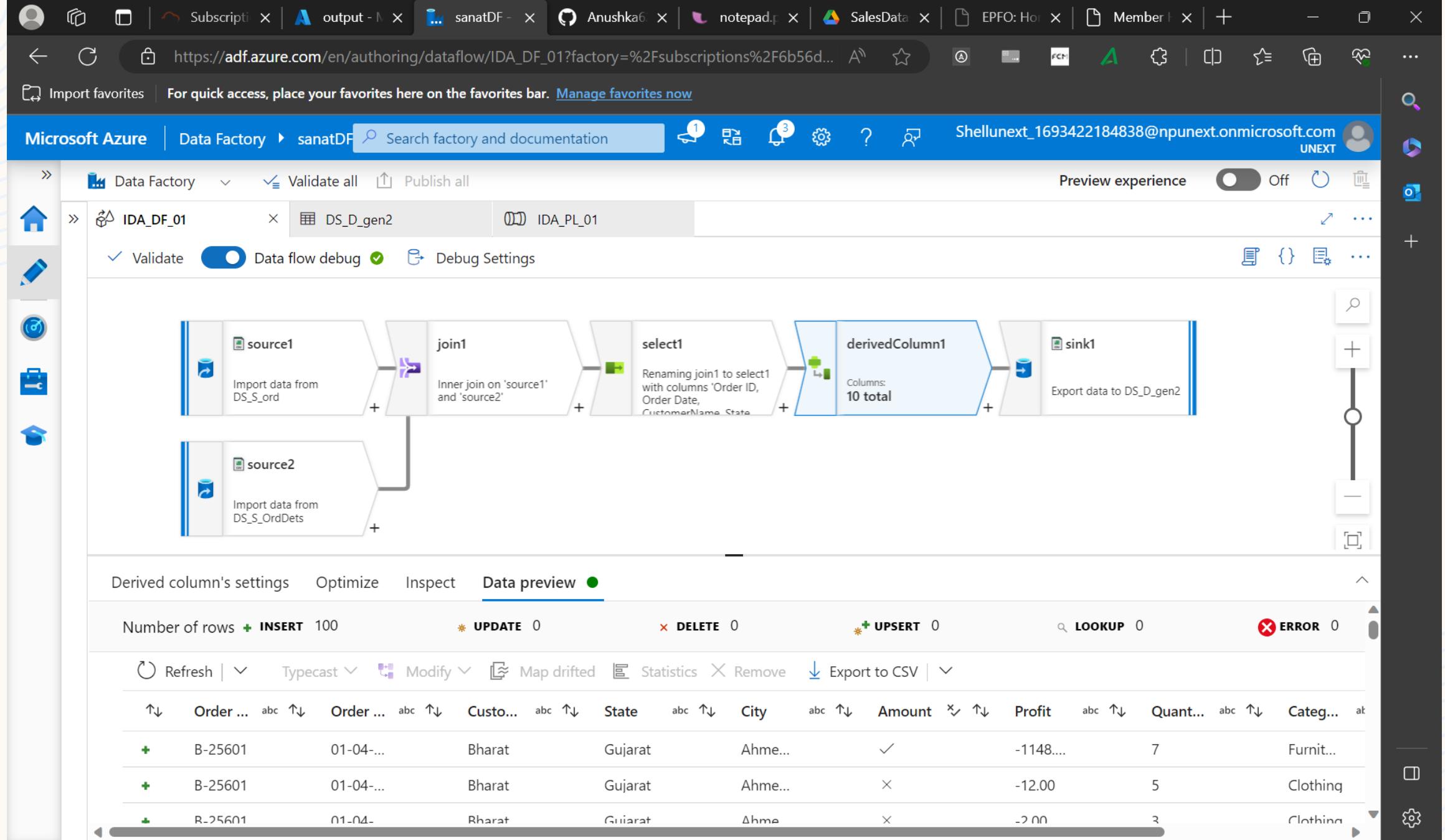
DAY 4

- Data(Resource)Transformation using data flow.
- For Each activity
- Data Flows
- Dataflow expression builder and derived columns
- Surrogate Key
- For Each container and Get Metadata
- API

 Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!







Subscription Det... | A output - Microsoft Edge | sanatDF - Azure | EPFO: Home | Member Home | dummyjson.com/ | +

Import favorites | For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure | Data Factory > sanatDF

Search factory and documentation

Publish all 1

Preview experience Off

Shellunext_1693422184838@npunext.onmicrosoft.com UNEXT

Factory Resources

Filter resources by name

Pipelines

- IDA_PL_01
- IDA_PL_02
- pipeline_api

Change Data Capture (preview) 0

Datasets

- DS_Copy
- DS_D_gen2
- DS_S_ord
- DS_S_OrdDets
- DS_source

Data flows

- IDA_DF_01

Power Query 0

Activities

Validate Debug Add trigger

Web

Web1

Output

Copy to clipboard

```
{ "todos": [ { "id": 1, "todo": "Do something nice for someone I care about", "completed": true, "userId": 26 } ] }
```

Pipeline status Succeeded View debug run consumption

Monitor in Azure Metrics Export to CSV

rt ↑	Duration ↑	Integration runtime	User properties ↑	A
Web1	Succeeded	9/7/2023, 3:18:22 PM	2s	AutoResolveIntegrator

3:19 PM 9/7/2023

DAY 5

- Create SQL Database on VM.
- Lookup Table in SQL Database
- Lookup table with ForEach pipeline
- CDC (Change Data Capture)
- Runbook or Automation

Subscription Details | N X | A containblob1 - Microso X sanatDF - Azure Data F X EPFO: Home Member Home X | +

Import favorites For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure | Data Factory > sanatDF

Search factory and documentation

Publish all 3

Preview experience Off

Factory Resources

Filter resources by name

Pipelines

- IDA_PL_01
- IDA_PL_02

Change Data Capture (preview) 0

Datasets

- DS_Copy
- DS_D_gen2
- DS_S_ord
- DS_S_OrdDets
- DS_source

Data flows

- IDA_DF_01

Power Query 0

Activities

Validate Debug Add trigger

IDA_DF_01 DS_D_gen2 IDA_PL_01 IDA_PL_02 DS_Copy

Get Metadata1

ForEach1

Copy data1

ForEach

Get Metadata1

Copy data1

Copy data1

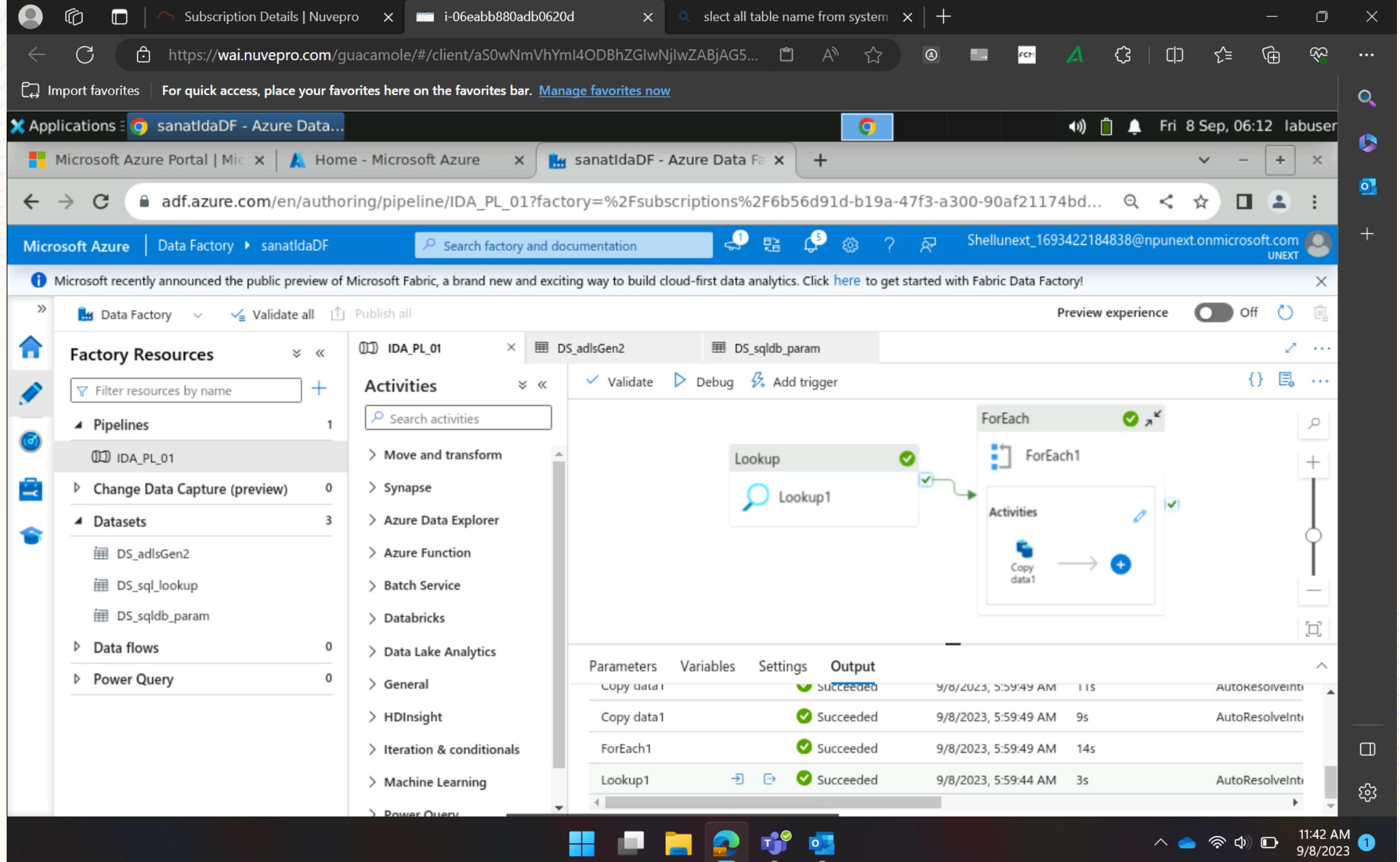
ForEach1

Get Metadata1

Output

Activity name	Activity status	Run start	Duration	Integration runtime	User properties
Copy data1	Succeeded	9/7/2023, 2:28:03 PM	9s	AutoResolveIntegrator	
Copy data1	Succeeded	9/7/2023, 2:28:03 PM	8s	AutoResolveIntegrator	
Copy data1	Succeeded	9/7/2023, 2:28:03 PM	8s	AutoResolveIntegrator	
ForEach1	Succeeded	9/7/2023, 2:28:02 PM	12s		
Get Metadata1	Succeeded	9/7/2023, 2:27:59 PM	3s	AutoResolveIntegrator	





Subscription Details | Nuvepro X i-06eabb880adb0620d X

Import favorites For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Applications sanatIldaDF - Azure Data... X

Microsoft Azure Portal | Mic x output - Microsoft Azure x sanatIldaDF - Azure Data Fa x +

adf.azure.com/en/authoring/adfcdc/CDCTest?factory=%2Fsubscriptions%2F6b56d91d-b19a-47f3-a300-90af21174bd2%2...

Microsoft Azure | Data Factory > sanatIldaDF Search factory and documentation

Shellunext_1693422184838@npunext.onmicrosoft.com UNEXT

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Data Factory Validate all Publish all Preview experience Off

Factory Resources Filter resources by name +

Pipelines IDA_PL_01 DS_adlsGen2 DS_sqldb_param CDCtest

IDCtest Start Set Latency Data flow debug ✓

This preview feature is licensed to you as part of your [Azure subscription](#). By continuing you agree to the [Preview Terms](#) and [Privacy Statement](#).

Your insights matter! Participate in our [brief survey](#) about our CDC top-level resource, and help us enhance your experience.

LS_adlsGen2

New mapping Delete

Source Table ↑ Target Table ↑

SalesLT.Customer output

Columns

Source Target

2:31 PM 9/8/2023

JOURNEY SCRAPBOOK

Custom Bootcamp – Week 3

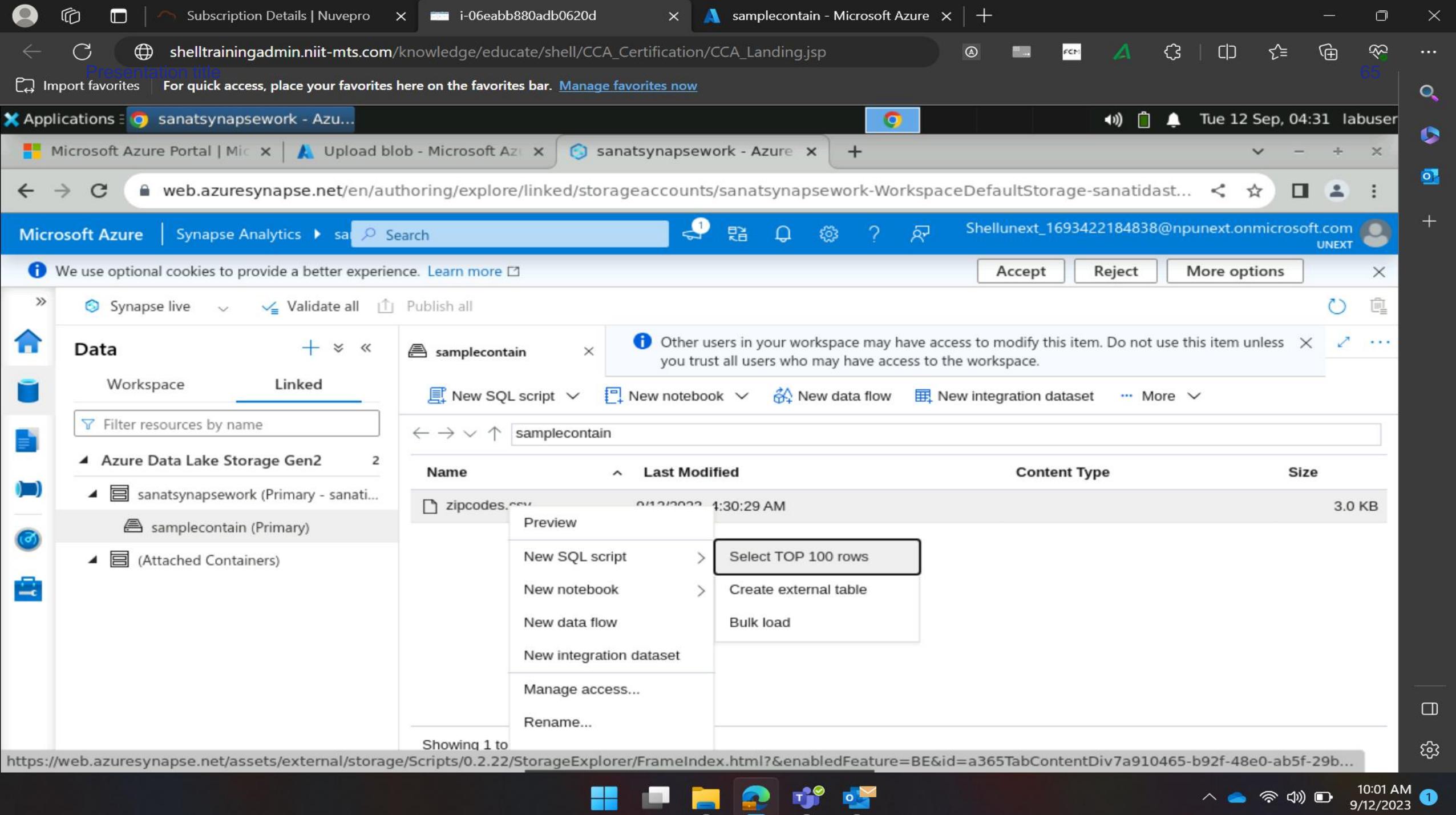
12 September, 2023 – 15 September, 2023

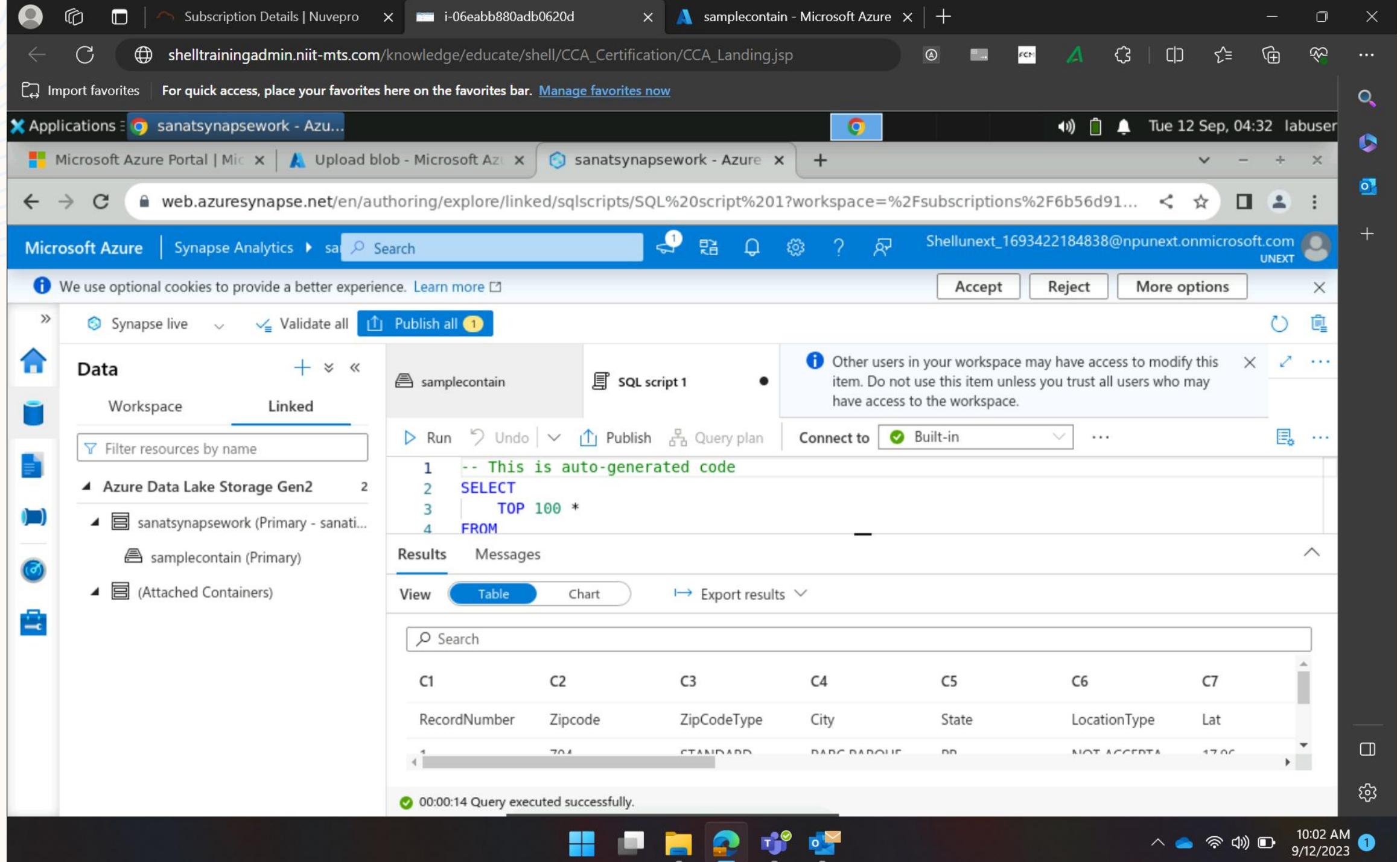
Sanat Mahajan

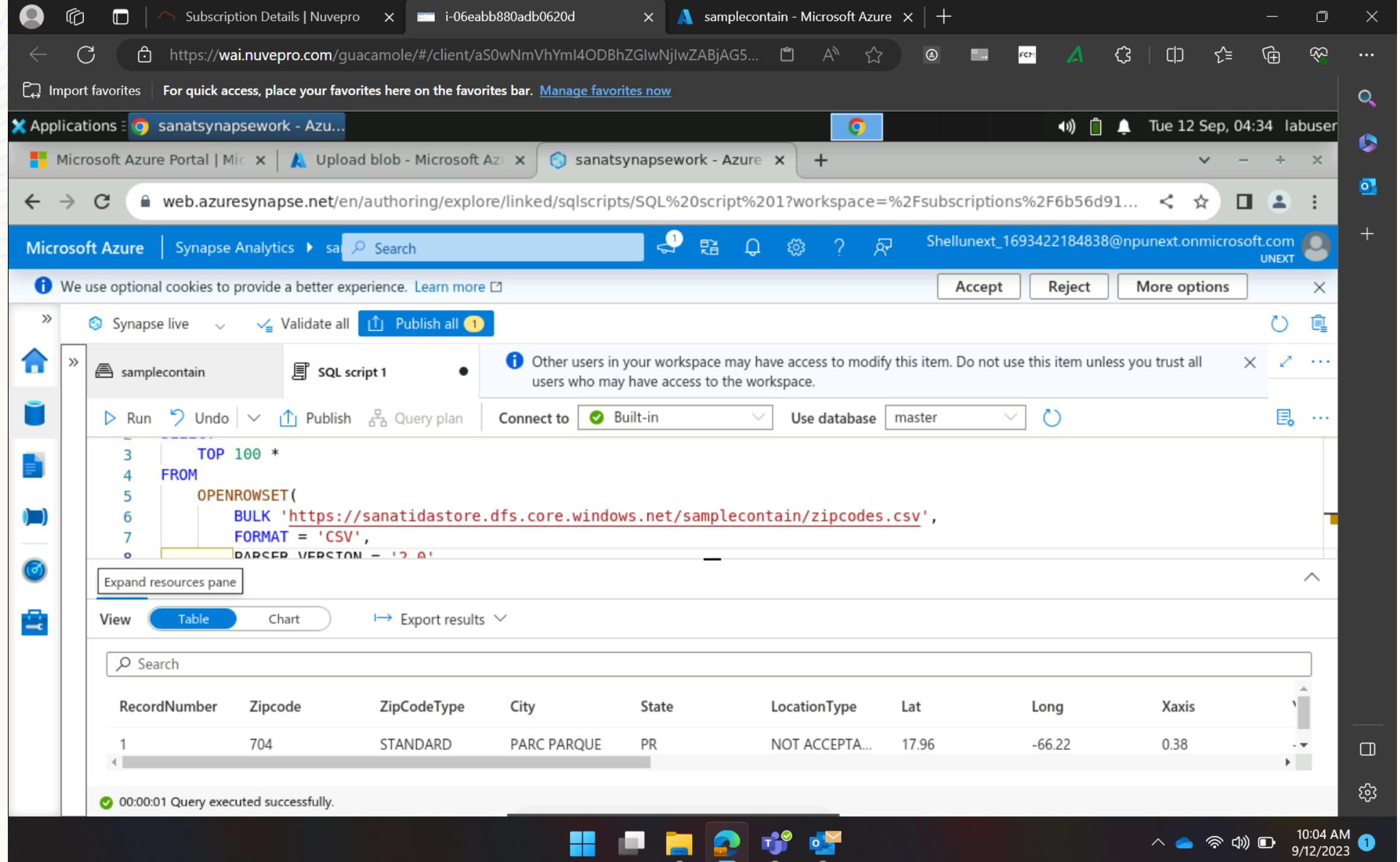
Batch 3

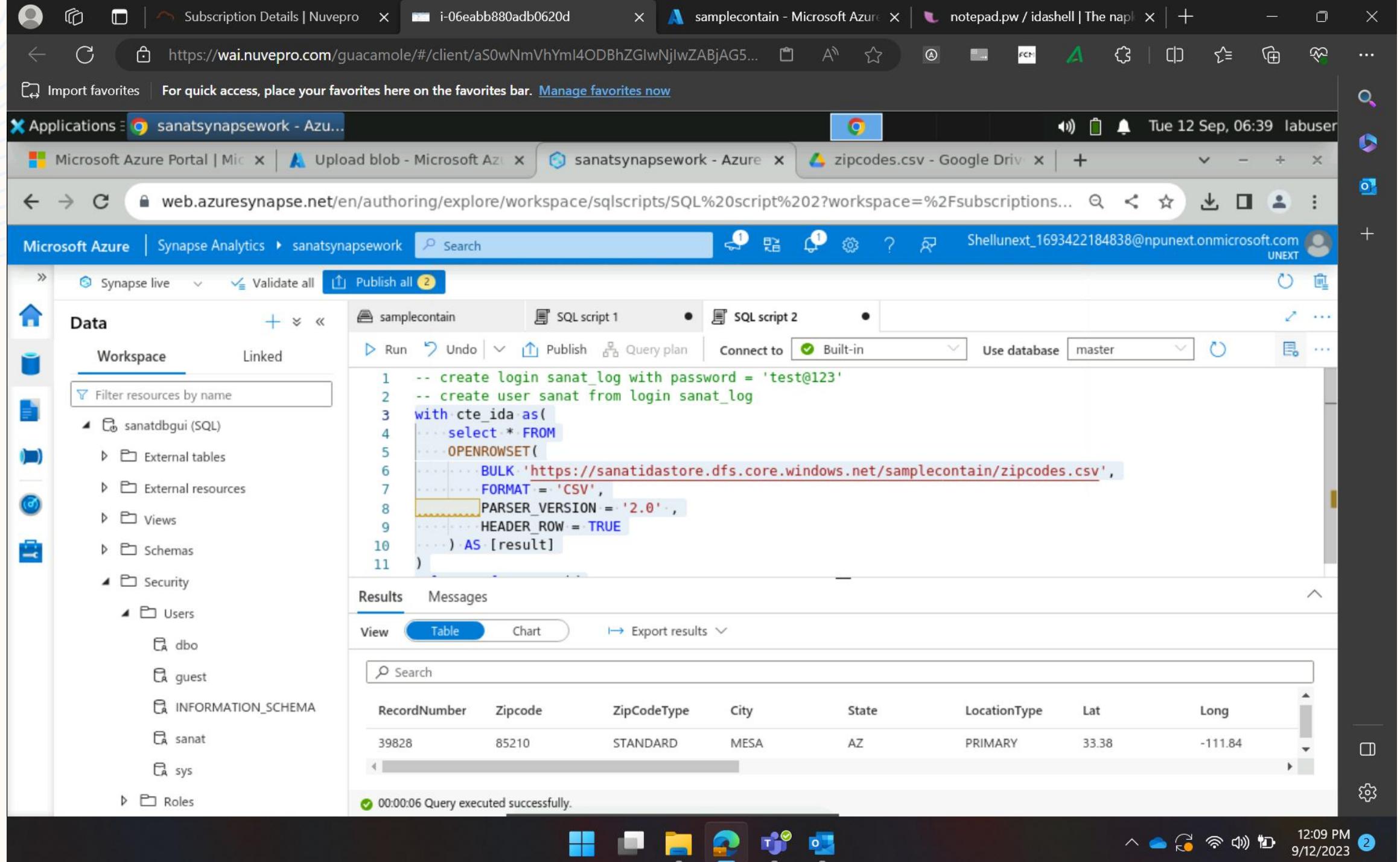
DAY 1

- Azure Synapse Analytics : Used to create ADF services, spark notebooks, SQL Data Warehouse
- SQL Pool
 - Serverless
 - Dedicated Pool
- Spark Pool
- Data Exploration
- SQL Database
- Collate
- Common Table Expression (CTE)
- View
- DWU (Data Warehousing Unit)
- Storage node
- Compute node
- Data Movement Service (DMS)
- Massive parallel processing
- Replicated Tables
- Hash Distributed Table
- Implementing Partitions for an SQL Data Warehouse



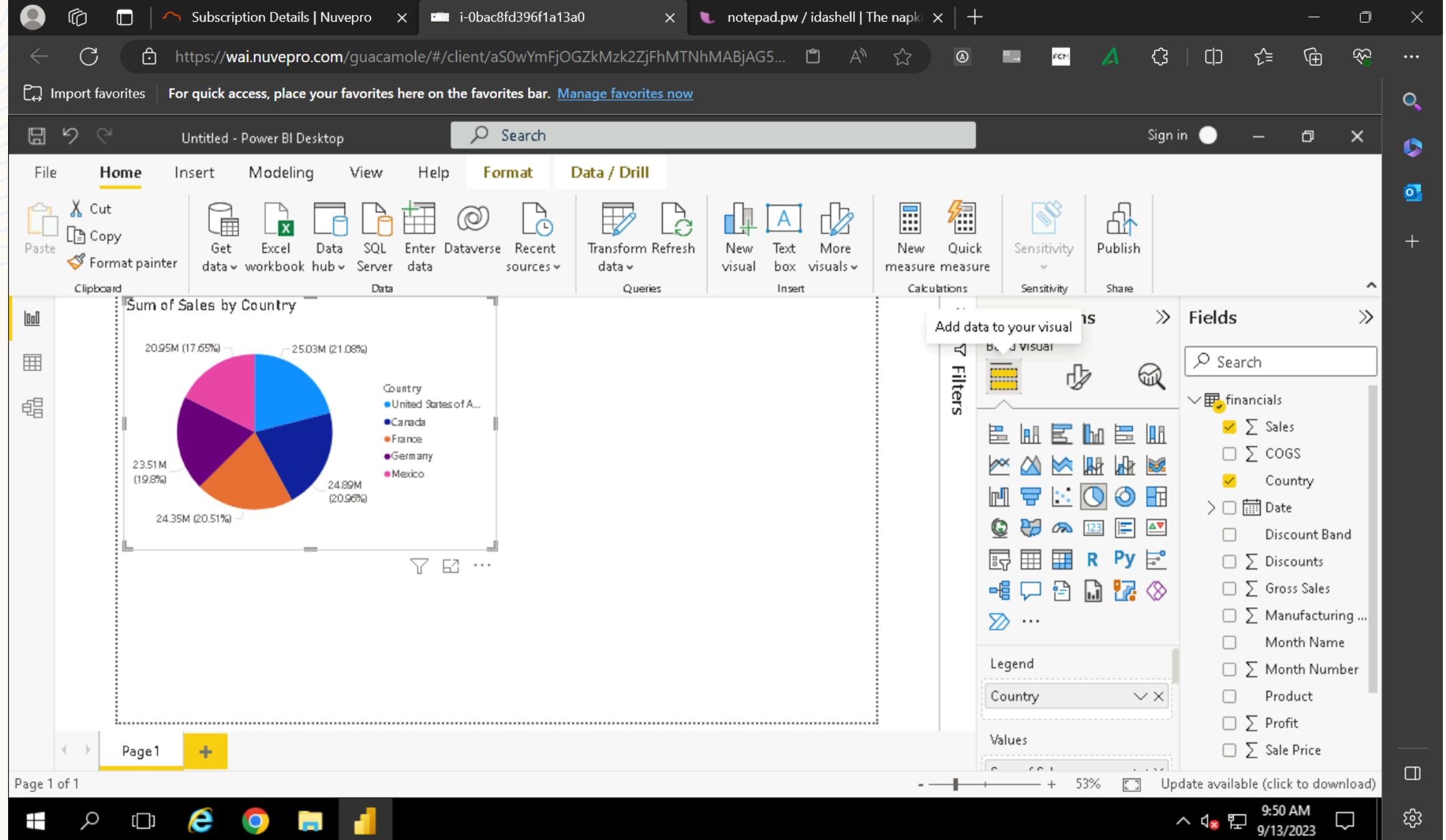






DAY 2

- PBI Desktop
- Power BI supports multiple data connectors.
- Power Query Editor: used to transform data
 - Language used at back-end - M Language
 - Different fields of this editor
- Get data in PBI report and transform it in Power Query Editor
- Getting familiar with different Visuals
- Parameter
- Filters :
 - Visual level
 - Page level
 - Report level
- Slicers
- DAX Measures
- DAX calculated columns
- RLS (Manage roles)



9:50 AM
9/13/2023

Subscription Details | Nuvepro X i-0bac8fd396f1a13a0 X truncate in power query - Search X 3 Ways to Truncate Numbers (Re X +

Import favorites For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Column From Examples Custom Invoke Function Conditional Column Index Column Duplicate Column Merge Columns ABC Extract Statistics Standard Scientific Trigonometry Rounding Date Time Duration Information Text Vision Azure Machine Learning AI Insights

General From Text From Number From Date & Time

Queries [3] financials customer segment_tbl

= Table.AddColumn(#"Replaced Value", "No of sales", each Number.IntegerDivide([Gross Sales],

	Month Number	Month Name	Year	No of sales
1	1/1/2014	January	2014	1618
2	1/1/2014	January	2014	1321
3	6/1/2014	June	2014	2178
4	6/1/2014	June	2014	888
5	6/1/2014	June	2014	2470
6	12/1/2014	December	2014	1513
7	3/1/2014	March	2014	921
8	6/1/2014	June	2014	2518
9	6/1/2014	June	2014	1899
10	6/1/2014	June	2014	1545
11	6/1/2014	June	2014	2470
12	7/1/2014	July	2014	2665
13	8/1/2014	August	2014	958
14	9/1/2014	September	2014	2146
15	10/1/2013	October	2013	345
16	12/1/2014	December	2014	615
17	2/1/2014	February	2014	292
18				

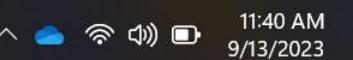
17 COLUMNS, 700 ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 9:46 AM

Query Settings

PROPERTIES

APPLIED STEPS

- Source
- Navigation
- Changed Type
- Replaced Value
- Added Custom



DAY 3

- Data modeling : establishing relationships between tables
 - Univariant
 - Bivariant
 - Multivariant
 - Drill Through
 - Date Table
 - Connection of SQL server and Power BI

pb_14-9 - Power BI Desktop

Search

Sign in

File Home Help

Cut Copy Paste Get data workbook hub Data SQL Server Enter data Dataverse Recent sources Transform Refresh data Manage relationships Manage roles View as Q&A setup Language setup Linguistic schema Sensitivity Publish Clipboard

Clipboard Data Queries Relationships Security Q&A Sensitivity Share

Auto recovery contains some recovered files that haven't been opened.

View recovered files

Properties

The screenshot shows the Power BI Data Model view with the following schema:

- Fact Tables:** part, partsupp, lineitem, orders.
- Dimension Tables:** supplier, customer, nation, region.
- Relationships:**
 - part (1) — partsupp (*).
 - partsupp (1) — lineitem (*).
 - lineitem (1) — orders (*).
 - supplier (1) — customer (*).
 - customer (1) — nation (*).
 - nation (1) — region (*).
- Attributes:** Each table lists various attributes such as P_BRAND, P_COMMENT, etc., for part, and S_ACCTBAL, C_ADDRESS, etc., for supplier.

3.110.84.173 - Remote Desktop Connection

pb_14-9 - Power BI Desktop

File Home Insert Modeling View Help

Cut Copy Format painter Clipboard

Get data workbook hub Data SQL Server Enter data Recent sources Data Transform Refresh data New visual Text box More visuals Insert New measure Quick measure Calculations Sensitivity Publish

Sensitivity Share

Region

- AFRICA
- AMERICA
- ASIA
- EUROPE
- MIDDLE EAST

Revenue Generated

A pie chart titled "Revenue Generated" showing revenue distribution across five regions. The data is as follows:

Region	Revenue	Percentage
MIDDLE EAST	386.17M	(18.15%)
AFRICA	469.34M	(22.06%)
ASIA	413....	(19....)
EUROPE	413.74M	(19.45%)
MIDDLE EAST	445.14M	(20.92%)

Nations

A world map titled "Nations" showing the location of various countries. The map includes labels for North America, South America, Europe, Africa, and Asia. It also shows the Atlantic and Indian Oceans.

Count of customers by nation

A bar chart titled "Count of customers by nation" showing the number of customers for 20 different nations. The Y-axis represents the count of customers, ranging from 0 to 80. The X-axis lists the nations. The chart shows a general decline in customer count from left to right.

Filters

Build visual

Values

Add data fields here

Drill through

Cross-report

Keep all filters

Add drill-through fields here

Visualizations Fields

Search

C_COMMENT C_CUSTKEY C_MKTSEGME C_NAME C_NATIONKE C_PHONE

DateTable

lineitem

nation

orders

O_CLERK O_COMMENT O_CUSTKEY

O_ORDERDAT O_ORDERKEY O_ORDERPRIC O_ORDERSTA

$\sum O_SHIPPRIOR$

$\sum O_TOTALPRIC$

2:16 PM 9/14/2023

3.110.84.173 - Remote Desktop Connection

pb_14-9 - Power BI Desktop

Search

Sign in

File Home Help Table tools

Name DateTable

Mark as date table Calendars Manage relationships New measure Quick New measure column New table Calculations

Structure

Data

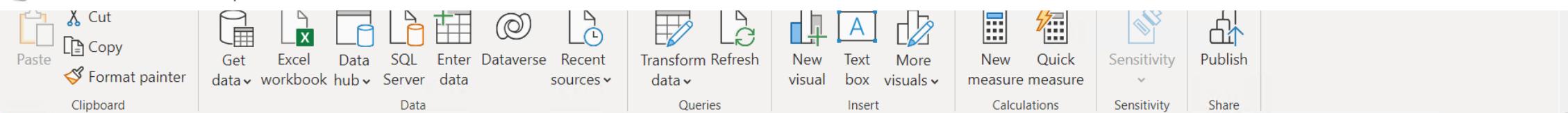
```
1 DateTable =  
2 ADDCOLUMNS(  
3     CALENDAR(DATE(2023,1,1), DATE(2023,12,31)),  
4     "Year", YEAR([Date]),  
5     "Month Number", MONTH([Date]),  
6     "Month Name", FORMAT([Date], "MMMM"),  
7     "Quarter", QUARTER([Date]),  
8     "Day of Week", WEEKDAY([Date]),  
9     "Day Name", FORMAT([Date], "dddd"),  
10    "Year-Month", FORMAT([Date], "YYYY-MM")  
11 )
```

Date	Year	Month Number	Month Name	Quarter	Day of Week	Day Name	Year-Month
7/1/2023 12:00:00 AM	2023	7	July	3	7	Saturday	2023-07
7/2/2023 12:00:00 AM	2023	7	July	3	1	Sunday	2023-07
7/3/2023 12:00:00 AM	2023	7	July	3	2	Monday	2023-07
7/4/2023 12:00:00 AM	2023	7	July	3	3	Tuesday	2023-07
7/5/2023 12:00:00 AM	2023	7	July	3	4	Wednesday	2023-07
7/6/2023 12:00:00 AM	2023	7	July	3	5	Thursday	2023-07
7/7/2023 12:00:00 AM	2023	7	July	3	6	Friday	2023-07
7/8/2023 12:00:00 AM	2023	7	July	3	7	Saturday	2023-07
7/9/2023 12:00:00 AM	2023	7	July	3	1	Sunday	2023-07
7/10/2023 12:00:00 AM	2023	7	July	3	2	Monday	2023-07
7/11/2023 12:00:00 AM	2023	7	July	3	3	Tuesday	2023-07
7/12/2023 12:00:00 AM	2023	7	July	3	4	Wednesday	2023-07
7/13/2023 12:00:00 AM	2023	7	July	3	5	Thursday	2023-07
7/14/2023 12:00:00 AM	2023	7	July	3	6	Friday	2023-07
7/15/2023 12:00:00 AM	2023	7	July	3	7	Saturday	2023-07

Fields

customer
DataTable
lineitem
nation
orders
part
partsupp
region
supplier

2:21 PM 9/14/2023 2



PetroCorp Report

Rig Name

Rig A

Rig B

Date

9/1/2023

9/5/2023

Total Drill Bit changes

7

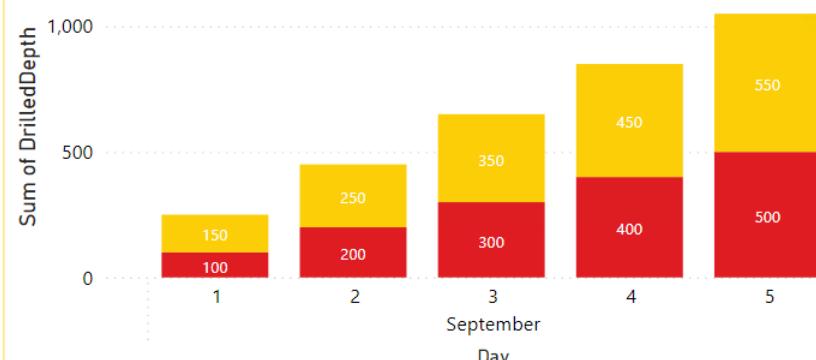
Total Drilling Speed

225.00

0.00 450.00

Total Drill Depth for the day

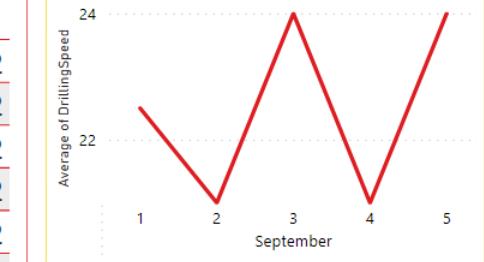
RigName • Rig A • Rig B



DataID | RigID

DataID	RigID
10	2
9	2
8	2
7	2
6	2
5	1
4	1
3	1
2	1
1	1

Average of Drilling Speed



Total Operating Hours

95

Page 1

+

Page 1 of 1

74%

Update available (click to down)

4:33 PM



4:33 PM

9/14/2023

2

DAY 4

1.Hands-on Assessment - Power BI

2.Jupyter Notebook launch

3.Datatypes in Python:

1. Integer
2. Float
3. String
4. List
5. Dictionary
6. Tuple
7. Boolean

3.Arithmetic

4.Logic



File

Home

Insert Design Layout References

Mailings

Review

View

Help

SHELL WIZKIT



Calibri (Body)

14

A⁺A⁻Aa⁺Aa⁻A^oA^z

Clipboard



Font

Paragraph

Styles

Editing

Voice

Sensitivity

Editor

Reuse Files

Name – Sanat Mahajan

Batch -3

Power Bi Hands On Assessment

Data Loading:

```
-- Create database
-- CREATE DATABASE PowerBIHandsOn;

-- Use Database
USE PowerBIHandsOn;

-- Create the Sales Transactions Data table
CREATE TABLE SalesTransactions (
    TransactionID INT PRIMARY KEY,
    ProductID INT,
    StoreID INT,
    TransactionDate DATE,
    QuantitySold INT,
    Revenue DECIMAL(10, 2)
);

-- Insert sample data into Sales Transactions Data
```



File Home Insert Modeling View Help

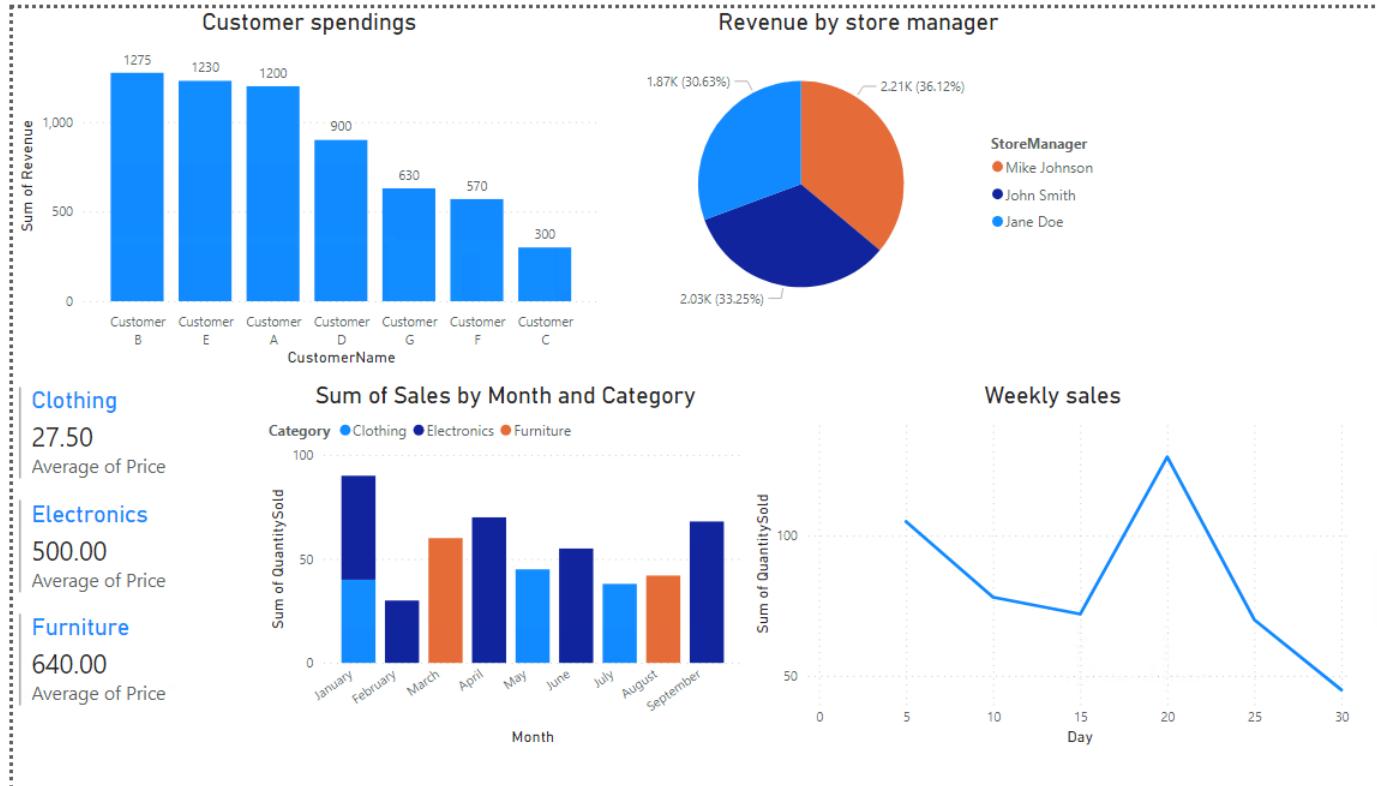
Paste Cut
Copy
Format painter
Clipboard

Get data v Excel Data SQL Server Enter data Dataverse Recent sources v

Transform data v Refresh data v New visual Text box More visuals v

New measure Quick measure Calculations

Sensitivity Publish
Sensitivity Share



Visualizations

Build visual

Filters

Values

Add data fields here

Drill through

Cross-report

Keep all filters

Add drill-through fields here

Fields

Search

- ProductName
- SalesTransactions
 - CustomerID
 - ProductID
 - QuantitySold
 - Revenue
- StoreID
- TransactionDate
 - Date Hierarchy
 - Year
 - Quarter
 - Month
 - Day
 - TransactionID
- Stores
 - Location
 - StoreID
 - StoreManager
 - StoreName
- StoreName H

15.207.18.195 - Remote Desktop Connection

Applications : Untitled - Jupyter Noteb... labuser@ip-172-31-14-2... Fri 15 Sep, 09:15

Python Sets Desktop/ Untitled - Jupyter Notebook notepad.pw / idashell | The bool to string in python - G

localhost:8888/notebooks/Desktop/Untitled.ipynb?kernel_name=python3 Logout

Jupyter Untitled Last Checkpoint: 3 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

```
5 # Floating-Point Numbers
6 price = 19.99
7 print("Price is",price)
8
9 # Strings
10 name = "Alice"
11 message = 'Hello, World!'
12 print(name+' '+message)
13
14 # Boolean Values
15 is_python_fun = True
16 is_learning = False
17 print("Sanat is not cool:",is_learning)
18 print("Anushka is good:" + str(is_python_fun))
19
20 # Lists
21 numbers = [1, 2, 3, 4, 5]
22 fruits = ["apple", "banana", "cherry"]
23 print(numbers[1:-1:2])
24 print(fruits)
25
26 # Tuples
27 coordinates = (3, 4)
28 print("Coordinates are :",coordinates[0],"°N",coordinates[1],"°W")
29
30 # Dictionaries
31 person = {"name": "Alice", "age": 30, "city": "New York"}
32 grades = {"math": 92, "history": 85, "science": 88}
33
34 # Accessing elements in lists and dictionaries
35 first_number = numbers[0]
36 last_fruit = fruits[-1]
37 name = person["name"]
38 math_grade = grades["math"]
39 print(first_number,last_fruit,name,math_grade)
```

Age is 25
Price is 19.99
Alice Hello, World!
Sanat is not cool: False
Anushka is good:True
[2, 4]
['apple', 'banana', 'cherry']
Coordinates are : 3 °N 4 °W
1 cherry Alice 92

In [15]: 1 a = input("Enter a no.")
2 print("Yo! mate, you entered",a)



15.207.18.195 - Remote Desktop Connection

Applications : Untitled - Jupyter Noteb... labuser@ip-172-31-14-2... Fri 15 Sep, 09:15

Python Sets Desktop/ Untitled - Jupyter Notebook notepad.pw / idashell | The G bool to string in python - G +

localhost:8888/notebooks/Desktop/Untitled.ipynb?kernel_name=python3 Logout

Jupyter Untitled Last Checkpoint: 3 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

2.0
0
100000
True

Out[19]: True

In [23]:

```
1 sta = [['a','b','c','d','e'],
2         ['a','c','d','l','k'],
3         ['w','e','i','r','d'],
4         ['a','s','h','o','l']]
5
6 class parameters:
7     profit = False
8     low_price = False
9     user_satisfaction = False
10
11     def __init__(self,profit,low_price,user_satisfaction):
12         self.profit = profit
13         self.low_price = low_price
14         self.user_satisfaction = user_satisfaction
15
16     def setProfit(self, profit):
17         self.profit = profit
18
19     def setLowPrice(self, low_price):
20         self.low_price = low_price
21
22     def setUserSat(self,user_sat):
23         self.user_satisfaction = user_sat
24
25 N_reg = parameters(False,True,True)
26 S_reg = parameters(True,True,True)
27 E_reg = parameters(True,False,False)
28 W_reg = parameters(True,True,False)
29
30 Region = [N_reg, S_reg, E_reg, W_reg]
31 for i in range(len(Region)):
32     profitable = Region[i].profit and (Region[i].low_price or Region[i].user_satisfaction)
33     if profitable:
34         sta[i].append('n')
35
36 print(sta)
37
[['a', 'b', 'c', 'd', 'e'], ['a', 'c', 'd', 'l', 'k'], ['w', 'e', 'i', 'r', 'd'], ['a', 's', 'h', 'o', 'l'],
'n']]
```

In []: 1



JOURNEY SCRAPBOOK

Custom Bootcamp – Week 4

19 September, 2023 – 22 September, 2023

Sanat Mahajan

Batch 3

DAY 1

1. Function
2. Function Composition
3. Recursive Function
4. Generator Function
5. Decorator (Wrap Function)
6. Types Of Argument
7. Class
8. Date Stripping
9. Date Conversion
10. Exception Handling
11. List Comprehension
12. Lambda Function

15.206.189.221 - Remote Desktop Connection

Applications : labuser@ip-172-31-14-2... Untitled - Jupyter Note...

Fri 22 Sep, 09:22

Desktop/ | pandas_tut - Jupyter N | Untitled1 - Jupyter Note | Untitled - Jupyter Note | PySpark_tut - Jupyter N | notepad.pw / idashell | Retail_Dataset - Google Sheets | python - Pyspark: disp

localhost:8888/notebooks/Desktop/Untitled.ipynb

jupyter Untitled Last Checkpoint: 09/15/2023 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted Python 3 (ipykernel)

In [24]:

```
1 # Integers
2 age = 25
3 print("Age is",age)
4
5 # Floating-Point Numbers
6 price = 19.99
7 print("Price is",price)
8
9 # Strings
10 name = "Alice"
11 message = 'Hello, World!'
12 print(name+' '+message)
13
14 # Boolean Values
15 is_python_fun = True
16 is_learning = False
17 print("Sanat is not cool:",is_learning)
18 print("Anushka is good:" + str(is_python_fun))
19
20 # Lists
21 numbers = [1, 2, 3, 4, 5]
22 fruits = ["apple", "banana", "cherry"]
23 print(numbers[1:-1:2])
24 print(fruits)
25
26 # Tuples
27 coordinates = (3, 4)
28 print("Coordinates are :",coordinates[0],"°N",coordinates[1],"°W")
29
30 # Dictionaries
31 person = {"name": "Alice", "age": 30, "city": "New York"}
32 grades = {"math": 92, "history": 85, "science": 88}
33
34 # Accessing elements in lists and dictionaries
35 first_number = numbers[0]
```

2:52 PM 9/22/2023

15.206.189.221 - Remote Desktop Connection

Applications : labuser@ip-172-31-14-2... Untitled - Jupyter Note... Fri 22 Sep, 09:23

Desktop/ | pandas_tut - Jupyter N... | Untitled1 - Jupyter Note... | Untitled - Jupyter Note... | PySpark_tut - Jupyter Note... | notepad.pw / idashell | Retail_Dataset - Google... | python - Pyspark: disp

localhost:8888/notebooks/Desktop/Untitled.ipynb

jupyter Untitled Last Checkpoint: 09/15/2023 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

```
7 profit = False
8 low_price = False
9 user_satisfaction = False
10
11 def __init__(self,profit,low_price,user_satisfaction):
12     self.profit = profit
13     self.low_price = low_price
14     self.user_satisfaction = user_satisfaction
15
16 def setProfit(self, profit):
17     self.profit = profit
18
19 def setLowPrice(self, low_price):
20     self.low_price = low_price
21
22 def setUserSat(self,user_sat):
23     self.user_satisfaction = user_sat
24
25 N_reg = parameters(False,True,True)
26 S_reg = parameters(True,True,True)
27 E_reg = parameters(True,False,False)
28 W_reg = parameters(True,True,False)
29
30 Region = [N_reg, S_reg, E_reg, W_reg]
31 for i in range(len(Region)):
32     profitable = Region[i].profit and (Region[i].low_price or Region[i].user_satisfaction)
33     if profitable:
34         sta[i].append('\n')
35
36 print(sta)
37
[['a', 'b', 'c', 'd', 'e'], ['a', 'c', 'd', 'l', 'k', 'n'], ['w', 'e', 'i', 'r', 'd'], ['a', 's', 'h', 'o', 'l', 'n']]
```

In []: 1



DAY 2

1. Pandas : Panel Datasets

- a. Dataframe
- b. Functions on dataframe
- c. Changing column to lowercase
- d. Dropping Null Values
- e. Dropping columns with Null Values
- f. Filling null values
- g. Describe columns
- h. Value counts()
- i. Selected rows or columns of dataframe
- j. Conditional selection

2. Matplotlib

3. Arrays In Python

15.206.189.221 - Remote Desktop Connection

Applications : labuser@ip-172-31-14-2... pandas_tut - Jupyter Note... Fri 22 Sep, 09:26

Desktop/ pandas_tut - Jupyter Note Untitled1 - Jupyter Note Untitled - Jupyter Note PySpark_tut - Jupyter Note notepad.pw / idashell Retail_Dataset - Google Sheets python - Pyspark: disp

jupyter pandas_tut Last Checkpoint: Last Wednesday at 10:41 AM (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel) Logout

In [1]: 1 import pandas as pd

In [2]: 1 data = {'orange': ['kashmir', 'ooty', 'bir'], 'apples': ['chennai', 'delhi', 'bglr']}
2 data

Out[2]: {'orange': ['kashmir', 'ooty', 'bir'], 'apples': ['chennai', 'delhi', 'bglr']}

In [3]: 1 fruit_df = pd.DataFrame(data, index = ['jan', 'feb', 'mar'])
2 fruit_df

Out[3]:

	orange	apples
jan	kashmir	chennai
feb	ooty	delhi
mar	bir	bglr

In [4]: 1 test = fruit_df.loc['feb']
2 test

Out[4]: orange ooty
apples delhi
Name: feb, dtype: object

In [5]: 1 movieDf = pd.read_csv("/home/labuser/Downloads/Pandas_datasets/IMDB-Movie-Data.csv", index_col = 0)
2 movieDf.head()

Out[5]:

Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
1	Guardians of the Galaxy	Action, Adventure, Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121	8.1	757074	333.13	76.0



15.206.189.221 - Remote Desktop Connection

Applications : labuser@ip-172-31-14-2... pandas_tut - Jupyter Not... Desktop/ x Untitled1 - Jupyter Note x Untitled - Jupyter Note x PySpark_tut - Jupyter Note x notepad.pw / idashell | x Google Retail_Dataset - Google Sheets x python - Pyspark: disp x + - x

Fri 22 Sep, 09:26 labuser

localhost:8888/notebooks/Desktop/pandas_tut.ipynb

jupyter pandas_tut Last Checkpoint: Last Wednesday at 10:41 AM (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Not Trusted Python 3 (ipykernel)

In [19]: `1 finalDf.isnull().sum()`

Out[19]:

	title	genre	description	director	actors	year	runtime	rating	votes	revenue_millions	metascore
dtype:	object										

In [19]: `1 finalDf.isnull().sum()`

Out[19]:

	title	genre	description	director	actors	year	runtime	rating	votes	revenue_millions	metascore
dtype:	int64										

In [20]: `1 finalDf.dropna()`

Out[20]:

Rank	title	genre	description	director	actors	year	runtime	rating	votes	revenue_millions	metascore
1	Guardians of the Galaxy	Action, Adventure, Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014	121	8.1	757074	333.13	76.0
2	Prometheus	Adventure, Mystery, Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012	124	7.0	485820	126.46	65.0
3	Split	Horror, Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016	117	7.3	157606	138.12	62.0
4	Sing	Animation, Comedy, Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdet	Matthew McConaughey, Reese Witherspoon, Seth Ma...	2016	108	7.2	60545	270.32	59.0
			A secret								



15.206.189.221 - Remote Desktop Connection

Applications: labuser@ip-172-31-14-2... pandas_tut - Jupyter Not... Desktop/ pandas_tut - Jupyter N... Untitled1 - Jupyter Note... Untitled - Jupyter Note... PySpark_tut - Jupyter N... notepad.pw / idashell | x Google Retail_Dataset - Google Sheets python - Pyspark: disp... Fri 22 Sep, 09:27 labuser

localhost:8888/notebooks/Desktop/pandas_tut.ipynb

jupyter pandas_tut Last Checkpoint: Last Wednesday at 10:41 AM (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Not Trusted Python 3 (ipykernel)

In [61]: 1 import matplotlib.pyplot as plt
Matplotlib is building the font cache; this may take a moment.

In [67]: 1 finalDf.plot(kind='scatter',x='rating',y='revenue_millions',title='Revenue vs Rating')

Out[67]: <Axes: title={'center': 'Revenue vs Rating'}, xlabel='rating', ylabel='revenue_millions'>

Revenue vs Rating

rating

revenue_millions



DAY 3

1. Driver process
2. Executors
3. Cluster Manager
4. Partition
5. Job, Stage and Task
6. Lazy Evaluation
7. RDD (Resilient Distributed Dataset)
8. Caching
9. Catalyst and Tungsten
10. RDD vs Dataframe
11. Deployment Types
 1. Client Mode
 2. Cluster Mode

15.206.189.221 - Remote Desktop Connection

Applications: labuser@ip-172-31-14-2... PySpark_tut - Jupyter N... Desktop/ PySpark_tut - Jupyter Note... notepad.pw / idashell | The... Retail_Dataset - Google Dr... python - Pyspark: display +

Fri 22 Sep, 09:29 labuser

localhost:8888/notebooks/Desktop/PySpark_tut.ipynb

jupyter PySpark_tut Last Checkpoint: 3 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [1]: 1 import findspark

In [2]: 1 findspark.init()

In [3]: 1 from pyspark.sql import SparkSession
2
3 spark = SparkSession.builder.appName('WordCount').getOrCreate()

Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/09/22 08:18:53 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable

In [4]: 1 sc = spark.sparkContext

In [5]: 1 rdd = sc.parallelize([1,2,3,4,5])

In [6]: 1 resultrdd = rdd.map(lambda x:x*2)
2 resultrdd.collect()

Out[6]: [2, 4, 6, 8, 10]

In [7]: 1 resultrdd = rdd.flatMap(lambda x:(x,x*2))
2 resultrdd.collect()

Out[7]: [1, 2, 2, 4, 3, 6, 4, 8, 5, 10]

In [8]: 1 resultrdd = rdd.filter(lambda x:x%2==0)
2 resultrdd.collect()

Out[8]: [2, 4]



15.206.189.221 - Remote Desktop Connection

Desktop/ PySpark_tut - Jupyter Notebooks/PySpark_tut.ipynb

jupyter PySpark_tut Last Checkpoint: 3 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Out[9]: [(2, 3), (1, 8), (3, 4)]

In [10]:

```
1 res2 = rdd1.groupByKey()
2
3 for key,values in res2.collect():
4     print(f"Key: {key}, Values: {list(values)}")
```

Key: 2, Values: [3]
Key: 1, Values: [2, 6]
Key: 3, Values: [4]

In [11]:

```
1 words_list = ['this', 'is', 'a', 'sample', 'text', 'document', 'for', 'word', 'count', 'exam', 'is', 'a', 'for']
```

In [12]:

```
1 rdd3 = sc.parallelize(words_list)
2 word_count = rdd3.flatMap(lambda x:x.split(','))
3 word_count.collect()
```

Out[12]: ['this',
'is',
'a',
'sample',
'text',
'document',
'for',
'word',
'count',
'exam',
'is',
'a',
'for']

In [13]:

```
1 wc = rdd3.map(lambda x:(x,1)).reduceByKey(lambda a,b:a+b)
2 wc.collect()
```

Out[13]: [('this', 1),
('sample', 1)]



15.206.189.221 - Remote Desktop Connection

Applications : labuser@ip-172-31-14-2... PySpark_tut - Jupyter N...

Desktop/ PySpark_tut - Jupyter Note... notepad.pw / idashell | The... Retail_Dataset - Google Dr... python - Pyspark: display ... +

Fri 22 Sep, 09:33 labuser

localhost:8888/notebooks/Desktop/PySpark_tut.ipynb

jupyter PySpark_tut Last Checkpoint: 3 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [13]:

```
1 wc = rdd3.map(lambda x:(x,1)).reduceByKey(lambda a,b:a+b)
2 wc.collect()
```

Out[13]:

```
[('this', 1),
 ('sample', 1),
 ('text', 1),
 ('for', 2),
 ('word', 1),
 ('is', 2),
 ('a', 2),
 ('document', 1),
 ('count', 1),
 ('exam', 1)]
```

In [14]:

```
1 purchaserdd = sc.textFile("/home/labuser/Downloads/Pandas_datasets/purchases.csv")
2 purchaserdd.collect()
```

Out[14]:

```
['', apples, oranges', 'June,3,0', 'Robert,2,3', 'Lily,0,7', 'David,1,2']
```

In [15]:

```
1 purchasedf = spark.read.csv("/home/labuser/Downloads/Pandas_datasets/purchases.csv")
2 purchasedf.show()
```

+---+---+---+

_c0	_c1	_c2
null	apples	oranges
June	3	0
Robert	2	3
Lily	0	7
David	1	2

+---+---+---+

In [16]:

```
1 df_01 = spark.read.option("inferSchema",True).option("header",True).csv("/home/labuser/Downloads/Pandas_datasets/...")
2 df_01.show()
3 df_01.printSchema()
```



DAY 4

1. Partitioning
2. Spark UI URL
3. SQL in PySpark
4. Drop duplicates
5. When – Otherwise condition
6. Concat and Functions
7. Caching
8. Joins
9. GroupBy

15.206.189.221 - Remote Desktop Connection

Desktop/ PySpark_tut - Jupyter Notebooks/PySpark_tut.ipynb

jupyter PySpark_tut Last Checkpoint: 3 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [21]: 1 testDf = moviesDf.repartition(10)

In [22]: 1 testDf.rdd.getNumPartitions()

23/09/22 08:19:28 WARN CSVHeaderChecker: Number of column in CSV header is not equal to number of fields in the schema:
Header length: 12, schema size: 3
CSV file: file:///home/labuser/Downloads/Pandas_datasets/IMDB-Movie-Data.csv

Out[22]: 10

In [24]: 1 testDf.write.csv("/home/labuser/Desktop/test")

23/09/22 08:21:21 WARN CSVHeaderChecker: Number of column in CSV header is not equal to number of fields in the schema:
Header length: 12, schema size: 3
CSV file: file:///home/labuser/Downloads/Pandas_datasets/IMDB-Movie-Data.csv

In [25]: 1 spark_ui_url = f'{spark._jsc.sc().uiWebUrl().get()}/'
2 print("Spark UI URL:",spark_ui_url)

Spark UI URL: http://ip-172-31-14-245.ap-south-1.compute.internal:4040/

In [26]: 1 sc.uiWebUrl

Out[26]: 'http://ip-172-31-14-245.ap-south-1.compute.internal:4040'

In [27]: 1 moviesDf.createOrReplaceTempView('movies')

In [28]: 1 res = spark.sql("select Rank,Title from movies")

In [29]: 1 type(res)



15.206.189.221 - Remote Desktop Connection

Desktop/ PySpark_tut - Jupyter Notebooks/PySpark_tut.ipynb

jupyter PySpark_tut Last Checkpoint: 3 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [31]: 1 from pyspark.sql.functions import *

In [32]: 1 moviesDf.sort(col("Title")).show()

Rank	Title	Genre
508	(500) Days of Summer	Comedy,Drama,Romance
119	10 Cloverfield Lane	Drama,Horror,Mystery
697	10 Years	Comedy,Drama,Romance
112	12 Years a Slave	Biography,Drama,H...
818	127 Hours	Adventure,Biograph...
169	13 Hours	Action,Drama,History
805	1408	Fantasy,Horror
923	17 Again	Comedy,Drama,Family
473	2012	Action,Adventure,...
248	20th Century Women	Comedy,Drama
851	21	Crime,Drama,Thriller
273	21 Jump Street	Action,Comedy,Crime
277	22 Jump Street	Action,Comedy,Crime
617	2307: Winter's Dream	Sci-Fi
876	28 Weeks Later	Drama,Horror,Sci-Fi
959	3 Days to Kill	Action,Drama,Thri...
431	3 Idiots	Comedy,Drama
114	300	Action,Fantasy,War
424	300: Rise of an E...	Action,Drama,Fantasy
777	31	Horror,Thriller

only showing top 20 rows

23/09/22 08:21:37 WARN CSVHeaderChecker: Number of column in CSV header is not equal to number of fields in the schema:
Header length: 12, schema size: 3
CSV file: file:///home/labuser/Downloads/Pandas_datasets/IMDB-Movie-Data.csv

In [33]: 1 moviesDf.select('Title').show()

3:05 PM 9/22/2023

15.206.189.221 - Remote Desktop Connection

Applications : labuser@ip-172-31-14-2... PySpark_tut - Jupyter N...

Desktop/ x PySpark_tut - Jupyter Note x notepad.pw / idashell | The x Retail_Dataset - Google Dr x python - Pyspark: display x +

→ C ① localhost:8888/notebooks/Desktop/PySpark_tut.ipynb

jupyter PySpark_tut Last Checkpoint: 3 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel) ○

Run Cell Kernel Cell Kernel Help Help

```
hema:
Header length: 12, schema size: 3
CSV file: file:///home/labuser/Downloads/Pandas_datasets/IMDB-Movie-Data.csv
```

In [40]: 1 movieDf.dropDuplicates(['Director']).count()

Out[40]: 646

In [41]: 1 movieDf.printSchema()

```
root
|-- Rank: integer (nullable = true)
|-- Title: string (nullable = true)
|-- Genre: string (nullable = true)
|-- Description: string (nullable = true)
|-- Director: string (nullable = true)
|-- Actors: string (nullable = true)
|-- Year: string (nullable = true)
|-- Runtime (Minutes): string (nullable = true)
|-- Rating: string (nullable = true)
|-- Votes: string (nullable = true)
|-- Revenue (Millions): double (nullable = true)
|-- Metascore: double (nullable = true)
```

In [42]: 1 ifDf = movieDf.withColumn("Rating_value",when((col("Rating")<=5) & (col("Rating")>2),"Average").when((col("Rating">5)),"High").otherwise("Low"))

In [43]: 1 ifDf.show()

Rank	Title	Genre	Description	Director	Actors
Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
1	Guardians of the ...	Action, Adventure, ...	A group of interg...	James Gunn	Chris Pratt, Vin ...



Applications : labuser@ip-172-31-14-2... PySpark_tut - Jupyter N... Fri 22 Sep, 09:37

Desktop/ PySpark_tut - Jupyter Note... notepad.pw / idashell | The... Retail_Dataset - Google Dr... python - Pyspark: display ... +

jupyter PySpark_tut Last Checkpoint: 3 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [53]: 1 new = cacDf.cache()

23/09/22 08:36:47 WARN CacheManager: Asked to cache already cached data.

In [55]: 1 new.show()

Rank	Title	Genre	Description	Director	Actors
Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
1	Guardians of the ...	Action, Adventure, ...	A group of interg...	James Gunn	Chris Pratt, Vin ...
2014	121	8.1	757074	333.13	76.0
2	Prometheus	Adventure, Mystery...	Following clues t...	Ridley Scott	Noomi Rapace, Log...
2012	124	7	485820	126.46	65.0
3	Split	Horror, Thriller	Three girls are k...	M. Night Shyamalan	James McAvoy, Any...
2016	117	7.3	157606	138.12	62.0
4	Sing	Animation, Comedy, ...	In a city of huma...	Christophe Lourdelet	Matthew McConaugh...
2016	108	7.2	60545	270.32	59.0
5	Suicide Squad	Action, Adventure, ...	A secret governme...	David Ayer	Will Smith, Jared...
2016	123	6.2	393727	325.02	40.0
6	The Great Wall	Action, Adventure, ...	European mercenar...	Yimou Zhang	Matt Damon, Tian ...
2016	103	6.1	56036	45.13	42.0
7	La La Land	Comedy, Drama, Music	A jazz pianist fa...	Damien Chazelle	Ryan Gosling, Emm...
2016	128	8.3	258682	151.06	93.0
8	Mindhorn	Comedy	"A has-been actor... whom he believes...	Sean Foley	Es sie Davis, Andr...
			2016	89	6.4
				2490.0	null

In [59]: 1 customerDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/I...
2 lineitemDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/I...
3 nationDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/D...
4 ordersDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/D...
5 partDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/D...
6 partsuppDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/I...
7 regionDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/D...
8 supplierDf = spark.read.option("sep", "\t").option("header", True).option("inferSchema", True).csv("/home/labuser/I...



JOURNEY SCRAPBOOK

Custom Bootcamp – Week 4

25 September, 2023 – 29 September, 2023

Sanat Mahajan

Batch 3

DAY 1

1. Caching
2. Persist and different storage levels
3. Aggregation
4. Managed and External Tables
5. Partitioning
6. For-Each

DAY 2

1. What is databricks?
2. Setting Up azure databricks
3. Overview of Databricks Workspace
4. Cluster configuration
5. Mounting
6. Widgets
7. Parameters
8. Functions
9. %run, %md, %language
10. Structured streaming

Subscription Details | Nuvepro X RG_sanatIDA_sanatDaBri - Microsoft Azure Create Cluster - Databricks X +

Import favorites For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure Search resources, services, and docs (G+/)

Shellunext_1693422184... UNEXT (NPUNEXT.ONMICROSOFT.COM)

Home > RG_sanatIDA_sanatDaBri | Overview X

Deployment

Search

Delete Cancel Redeploy Download Refresh

Overview Inputs Outputs Template

✓ Your deployment is complete

Deployment name : RG_sanatIDA_sanatDaBri Start time : 9/26/2023, 9:39:31 AM
Subscription : npunext-1673505552909 Correlation ID : 889696cf-ce78-4e5b-81f8-...
Resource group : RG_sanatIDA

> Deployment details

▽ Next steps

Go to resource

Give feedback

Tell us about your experience with deployment

Cost management Get notified to stay within your budget and prevent unexpected charges on your bill. Set up cost alerts >

Microsoft Defender for Cloud Secure your apps and infrastructure Go to Microsoft Defender for Cloud >

Free Microsoft tutorials Start learning today >

Work with an expert Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.

10:05 AM 9/26/2023

Subscription Details | Nuvepro X RG_sanatIDA_sanatDaBri - Microsoft Azure X Home - Workspace - Databricks X +

Import favorites For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure |  databricks Search data, notebooks, recents, and more... CTRL + P sanatDaBri shellunext_1693422184838@npunext.onmicrosoft.com

+ New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards NEW

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Delta Live Tables

Workspace

Home

Workspace

Repos

Trash

Workspace > Users > **shellunext_1693422184838@npunext.onmicrosoft.com**

Free trial ends in 14 days. [Upgrade to Premium in Azure Portal](#)

Share

Name ▲ Type Owner Created

Untitled Notebook 2023-09-26 0... Notebook Shellunext 9:44 AM

<https://adb-6969336054380226.azuredatabricks.net/?o=6969336054380226>



Subscription Details | Nuvepro X | RG_sanatIDA_sanatDaBri - Microsoft Azure X | Cluster Details - Databricks X +

Import favorites For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure |  databricks Search data, notebooks, recents, and more... CTRL + P sanatDaBri ? shellunext_1693422184838@npun...

+ New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards NEW Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning

Compute > UI preview Send feedback

Shellunext's Cluster

Configuration Notebooks (0) Libraries Event log Spark UI Driver logs Metrics Apps Spark compute UI - Master ▾

Single user Shellunext Photon Standard_DS3_U1 JSON /

Performance

Databricks Runtime Version
13.3 LTS (includes Apache Spark 3.4.1, Scala 2.12)

Use Photon Acceleration ?

Node type ?
Standard_DS3_v2 14 GB Memory, 4 Cores

Terminate after 20 minutes of inactivity ?

Tags ?

No custom tags

> Automatically added tags

> Advanced options

10:48 AM 9/26/2023

DAY 3

1. Structured streaming
2. DBFS from sample data
3. Delta table & Parquet table
4. Partition Tables
5. Medallion / Multi hop Architecture
6. Workflows → Jobs
7. Accessing data from Azure SQL database
8. Unity catalog in databricks
9. Metastore

Untitled Notebook 2023-09-26 09:44:59

Python ▾



File Edit View Run Help Last edit was 5 minutes ago Provide feedback

▶ Run all

● Shellunext's Cluster ▾

Schedule

Share

Cmd 1

Python ▶ ▾ - x

```
1 dbutils.fs.mount(source = "wasbs://input@sanatidagen2.blob.core.windows.net",mount_point = "/mnt/input",
extra_configs = {"fs.azure.account.key.sanatidagen2":dbutils.secrets.get(scope = "sanatDBscope", key =
"sanatKeygen2")})
```

⊕shaded.databricks.org.apache.hadoop.fs.azure.AzureException: shaded.databricks.org.apache.hadoop.fs.azure.AzureException: Unable to access container input in account sanatidagen2.blob.core.windows.net using anonymous credentials, and no credentials found for them in the configuration.

Command took 1.38 seconds -- by shellunext_1693422184838@npunext.onmicrosoft.com at 9/26/2023, 12:39:19 PM on Shellunext's Cluster

Cmd 2

Python ▶ ▾ - x

```
1
```

Shift+Enter to run

Shift+Ctrl+Enter to run selected text

Subscription | sanatDaBri - | Catalog Explorer | shell hub - S | Home | Home | :: Shell :: | - | X

Import favorites For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure | databricks Search data, notebooks, recents, and more... CTRL + P sanatDaBri ? shellunext_1693422184838@npun... +

+ New Workspace Recents Catalog Workflows Compute SQL SQL Editor Queries Dashboards NEW Alerts Query History SQL Warehouses Data Engineering Job Runs Data Ingestion Delta Live Tables Machine Learning

Catalog Explorer Provide Feedback

Type to filter

default >

default.zipcodes_1_csv ▲

Owner: Not set Size: 6.3KiB, 1 file Last Updated: Unknown

Comment: Add comment

+ Add Shellunext's Cluster 14 GB, 4 Cores Create

Columns Sample Data Details Permissions History

Filter columns...

Column ⚡	Type	Comment
RecordNumber	int	⊕
Zipcode	int	⊕
ZipCodeType	string	⊕
City	string	⊕
State	string	⊕
LocationType	string	⊕
Lat	double	⊕

9:49 AM 9/27/2023 8

Subscription Det... | input - Microsoft | saanatKeyVault - | 27-9 - Databricks | Onboarding | GitHub - Anushka... | +

Import favorites | For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Microsoft Azure | databricks | Search data, notebooks, recents, and more... CTRL + P sanatDaBri ? shellunext_1693422184838@npun...

+ New

Workspace

- Recents
- Catalog
- Workflows
- Compute
- SQL
- SQL Editor
- Queries
- Dashboards NEW
- Alerts
- Query History
- SQL Warehouses
- Data Engineering
- Job Runs
- Data Ingestion
- Delta Live Tables
- Machine Learning

27-9 Python ★

Last edit was 19 minutes ago Provide feedback

Run all Shellunext's Cluster Schedule Share

```
1 df = spark.read.option("inferschema", True).option("header", True).csv("/mnt/input/zipcodes.csv")
```

(2) Spark Jobs

df: pyspark.sql.dataframe.DataFrame = [RecordNumber: integer, Zipcode: integer ... 18 more fields]

Command took 0.98 seconds -- by shellunext_1693422184838@npunext.onmicrosoft.com at 9/27/2023, 11:23:26 AM on Shellunext's Cluster

Cmd 4

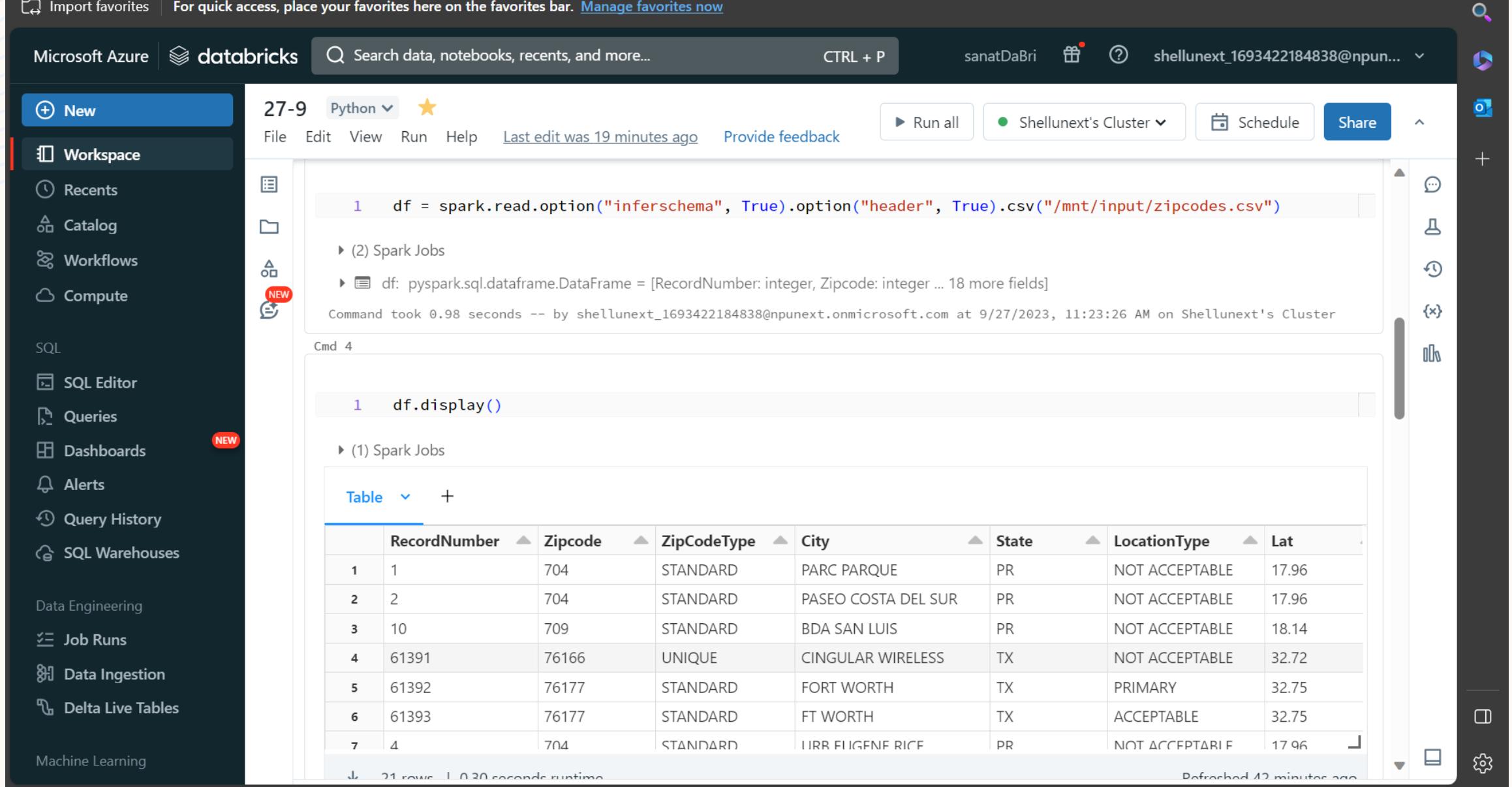
```
1 df.display()
```

(1) Spark Jobs

Table +

	RecordNumber	Zipcode	ZipCodeType	City	State	LocationType	Lat
1	1	704	STANDARD	PARC PARQUE	PR	NOT ACCEPTABLE	17.96
2	2	704	STANDARD	PASEO COSTA DEL SUR	PR	NOT ACCEPTABLE	17.96
3	10	709	STANDARD	BDA SAN LUIS	PR	NOT ACCEPTABLE	18.14
4	61391	76166	UNIQUE	CINGULAR WIRELESS	TX	NOT ACCEPTABLE	32.72
5	61392	76177	STANDARD	FORT WORTH	TX	PRIMARY	32.75
6	61393	76177	STANDARD	FT WORTH	TX	ACCEPTABLE	32.75
7	4	704	STANDARD	FLOR FLUGENF RICE	PR	NOT ACCEPTABLE	17.96

21 rows | 0.30 seconds runtime Refreshed 12 minutes ago



DAY 4

1. What is Docker ?
2. Docker initialization in VM
3. Build Docker image
 - Application
 - Requirement.txt
 - Dockerfile
4. Image created and Docker push
5. Azure container registry (ACR)
6. Kubernetes
7. AKS
8. Deploy a single-image application using code and UI

65.0.5.130 - Remote Desktop Connection

Applications : Firefox notepad.pw / idashell | ... root@ip-172-31-4-204: /...

root@ip-172-31-4-204: /home/labuser/Desktop/docker_project

File Edit View Search Terminal Tabs Help

root@ip-172-31-4-204: /home/labuser/Desktop/docker_project labuser@ip-172-31-4-204: ~/Desktop/docker_project

```
Step 2/7 : WORKDIR /app
--> Using cache
--> 899c2aca6e1f
Step 3/7 : COPY . /app
--> 6ba2d29fb30b
Step 4/7 : RUN pip install -r requirements.txt
--> Running in a362e455fa22
Collecting Flask==2.0.1
  Downloading Flask-2.0.1-py3-none-any.whl (94 kB) 94.8/94.8 kB 6.8 MB/s eta 0:00:00
Collecting Werkzeug>=2.0
  Downloading werkzeug-2.3.7-py3-none-any.whl (242 kB) 242.2/242.2 kB 18.8 MB/s eta 0:00:00
Collecting click>=7.1.2
  Downloading click-8.1.7-py3-none-any.whl (97 kB) 97.9/97.9 kB 12.0 MB/s eta 0:00:00
Collecting itsdangerous>=2.0
  Downloading itsdangerous-2.1.2-py3-none-any.whl (15 kB)
Collecting Jinja2>=3.0
  Downloading Jinja2-3.1.2-py3-none-any.whl (133 kB) 133.1/133.1 kB 8.3 MB/s eta 0:00:00
Collecting MarkupSafe>=2.0
  Downloading MarkupSafe-2.1.3-cp38-cp38-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (25 kB)
Installing collected packages: MarkupSafe, itsdangerous, click, Werkzeug, Jinja2, Flask
Successfully installed Flask-2.0.1 Jinja2-3.1.2 MarkupSafe-2.1.3 Werkzeug-2.3.7 click-8.1.7 itsdangerous-2.1.2
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: https://pip.pypa.io/warnings/venv
```

[notice] A new release of pip is available: 23.0.1 -> 23.2.1
[notice] To update, run: pip install --upgrade pip

Removing intermediate container a362e455fa22
--> 4eab4b201be9

Step 5/7 : EXPOSE 80
--> Running in f1546c22dd83

Removing intermediate container f1546c22dd83
--> c4b07c8eb7f1

Step 6/7 : ENV NAME World
--> Running in ab983e8f17b8

Removing intermediate container ab983e8f17b8
--> 3bf4741b0e8f

Step 7/7 : CMD ["python", "app.py"]
--> Running in 3a6bf0e753f1

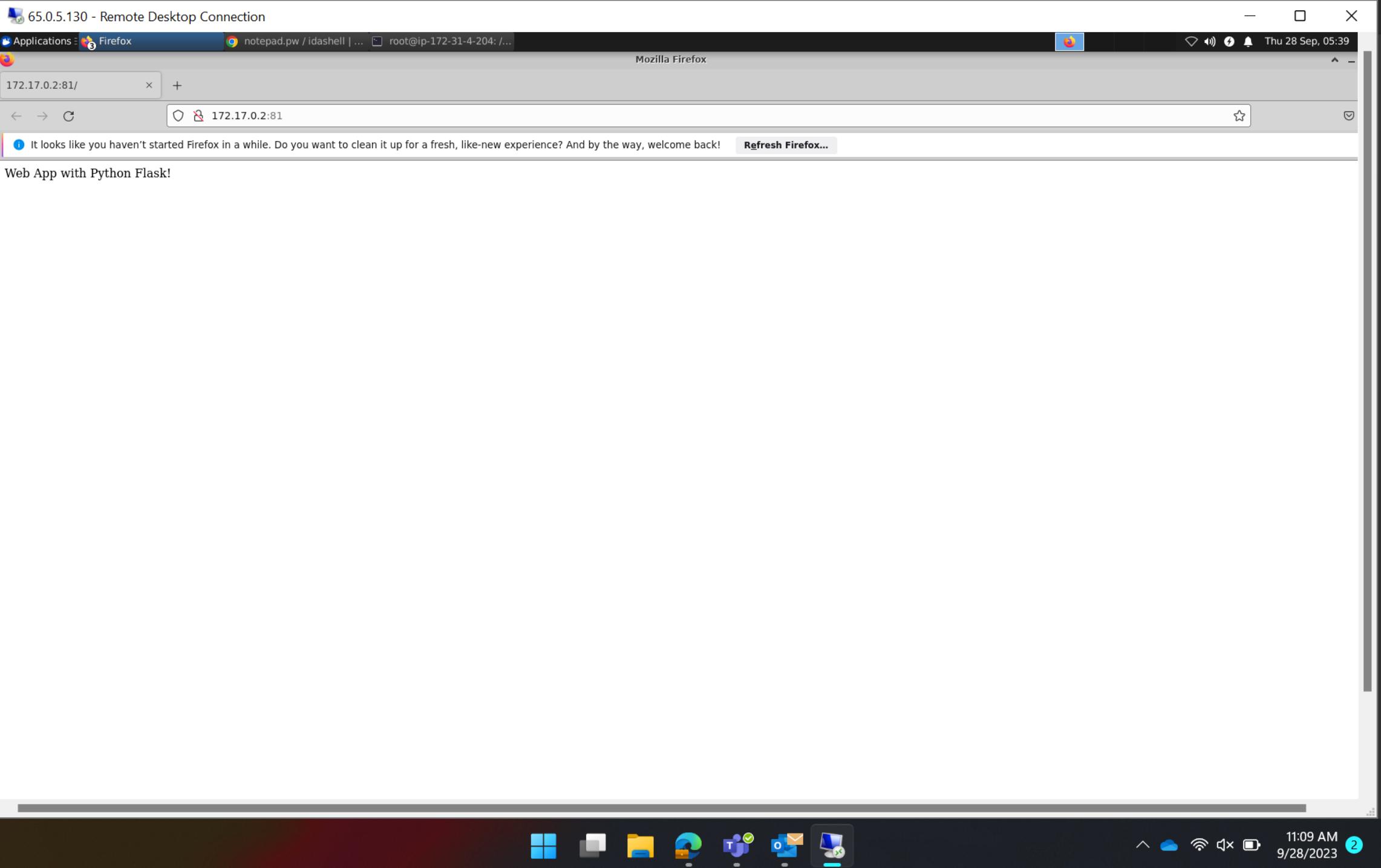
Removing intermediate container 3a6bf0e753f1
--> 19ae05e639c0

Successfully built 19ae05e639c0

Successfully tagged my-python-app:latest

```
root@ip-172-31-4-204:/home/labuser/Desktop/docker_project# docker images list
REPOSITORY TAG IMAGE ID CREATED SIZE
root@ip-172-31-4-204:/home/labuser/Desktop/docker_project# docker images list
REPOSITORY TAG IMAGE ID CREATED SIZE
root@ip-172-31-4-204:/home/labuser/Desktop/docker_project# docker run my-python-app
* Serving Flask app 'app' (lazy loading)
* Environment: production
WARNING: This is a development server. Do not use it in a production deployment.
Use a production WSGI server instead.
* Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
```

11:09 AM 9/28/2023



11:09 AM
9/28/2023

https://wai.nuvepro.com/guacamole/#/client/a50wYWQ3MWM2YWM1NDYyOTg0MwBj... labuser@ip-172-31-4-20... [kubernetes - File Manager] Chrome Thu 28 Sep, 10:50 labuser

Applications Deploy a quickstart app... labuser@ip-172-31-4-20... [kubernetes - File Manager] Microsoft Azure Portal | Mic Deploy a quickstart applica Azure Voting App notepad.pw / idashell | The ... + ← → C portal.azure.com/#view/Microsoft_Azure_ContainerService/K8QuickStartApplication.ReactView/aksClusterId/%2... Download Share Star Print Update : Microsoft Azure Search resources, services, and docs (G+/) Home > AnuKuberClust > Create a starter application > Deploy a quickstart application ...

Resource Type Status

✓ azure-vote	Namespace	Success
✓ azure-vote-back	Deployment	Success
✓ azure-vote-back	Service	Success
✓ azure-vote-front	Deployment	Success
✓ azure-vote-front	Service	Success

2. Next steps

Here are some actions you can take to further explore your new application and learn

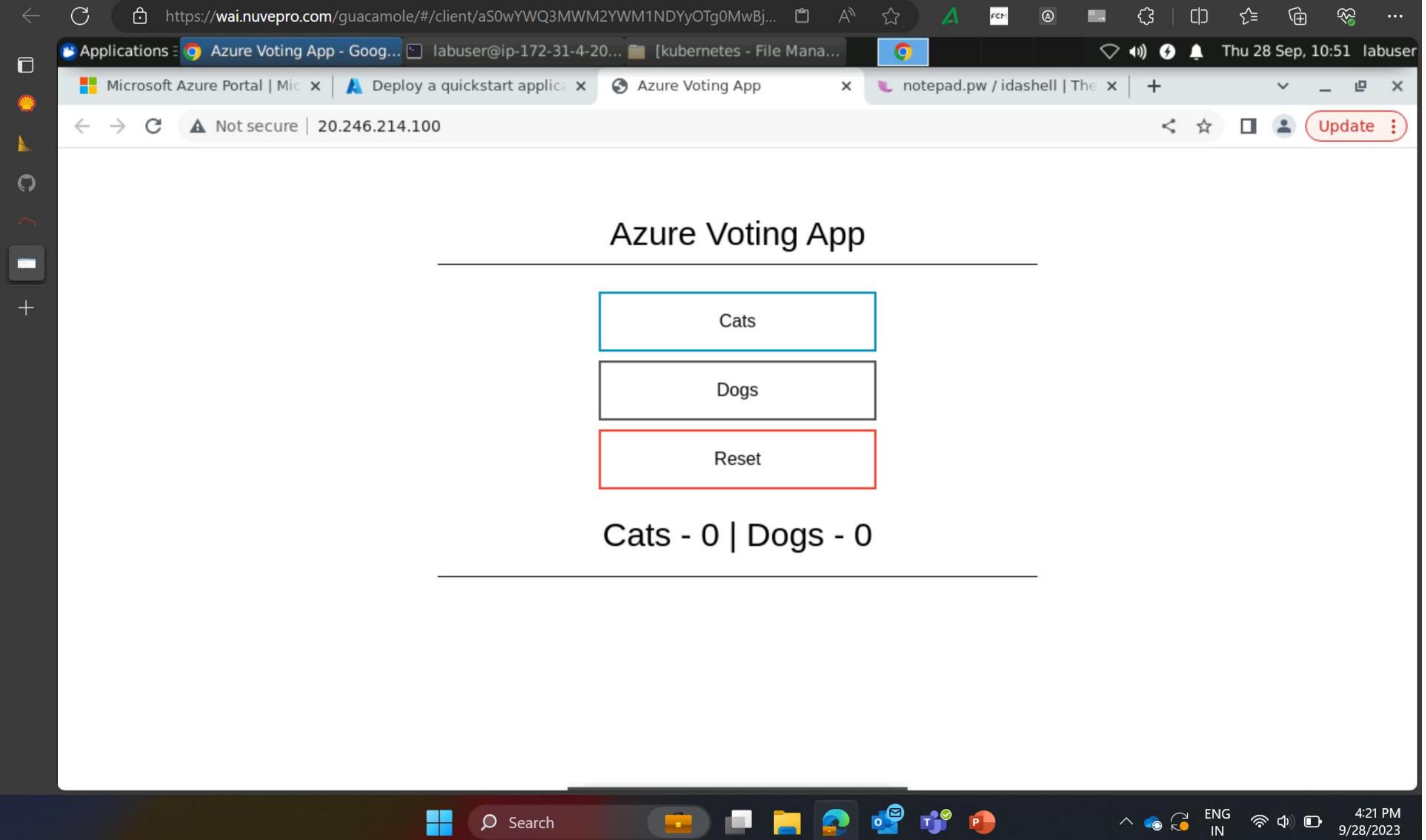
Previous Close



Search



4:20 PM
ENG
IN
9/28/2023



Azure Voting App

Cats

Dogs

Reset

Cats - 0 | Dogs - 0



Search



ENG
IN



4:21 PM
9/28/2023

DAY 5

1. What is DevOps ?
 - Collaborative Branch (Main branch)
 - Feature branch
2. Azure DevOps
 - Boards
 - Repos
 - Pipelines
 - Test plans
3. Creating a project Board in Azure DevOps
 - Work items
 - Epic
 - Feature
 - User story
 - Tasks
4. Azure Repos
5. Azure Pipelines

Azure DevOps Shellunext1693422200894 / IDA_Anushka_Shell / Boards / Work items

Search

ID A Anushka_Shell + Try the New Boards Hub for improved performance, accessibility, and new features. [Click here](#) to learn more.

Work items

Recently updated | + New Work Item | Open in Queries | Column Options | Import Work Items | Recycle Bin

Filter by keyword

ID	Title	Assigned To	State	Area Path	Tags
24	Implement a search algorithm that queries the database for books	Unassigned	New	IDA_Anushka_Shell	
23	Design a search bar and results page UI	Unassigned	New	IDA_Anushka_Shell	
22	As a User, I want to search for books by title or author, so i can ...	Unassigned	New	IDA_Anushka_Shell	
21	Implement backend logic to fetch books by category	Unassigned	New	IDA_Anushka_Shell	
20	Develop a UI for browsing books by category	Unassigned	New	IDA_Anushka_Shell	
19	Create a database schema for book categories	Unassigned	New	IDA_Anushka_Shell	
18	As a user, I want to browse books by category, so i can find books ...	Unassigned	New	IDA_Anushka_Shell	
17	Search books	Unassigned	New	IDA_Anushka_Shell	
16	Browse books	Unassigned	New	IDA_Anushka_Shell	
15	Browse and Search Books	Unassigned	New	IDA_Anushka_Shell	
14	Handle authentication errors and provide appropriate feedback	Unassigned	New	IDA_Anushka_Shell	
13	Implement backend logic for user authentication	Unassigned	New	IDA_Anushka_Shell	
12	Design the login page UI	Unassigned	New	IDA_Anushka_Shell	
11	As a registered user, i want to log in to my account	Unassigned	New	IDA_Anushka_Shell	
10	User Login	Unassigned	New	IDA Anushka Shell	

Boards Backlogs Sprints Queries Delivery Plans Analytics views Repos Pipelines Test Plans Artifacts Project settings

IDA_Anushka_Shell +

Overview

Boards

Repos

Files

Commits

Pushes

Branches

Tags

Pull requests

Advanced Security

Pipelines

Test Plans

Artifacts

Project settings

Branches

New branch

Mine All Stale

Search branch name

Branch	Com...	Author	Authored ...	Behind Ahead	Status	Pull ...
feature	21c4f62	 Shellunext unex...	52m ago	1 0		
main	b03970f	 Shellunext unex...	51m ago			

Default Compare

Azure DevOps Shellunext1693422200894 / IDA_Anushka_Shell / Repos / Pull requests / IDA_Anushka_Shell

Search

IDA_Anushka_Shell +

New pull request

Overview Boards Repos Files Commits Pushes Branches Tags Pull requests Advanced Security Pipelines Test Plans Artifacts

Pull requests

Mine Active Completed Abandoned

Pull Request ID	Created by	Assigned to	Target branch
SU updated test Shellunext unext!DA90 request !3 into 99 main	SU	0	Completed 52m ago
SU Updated test.json Shellunext unext!DA90 request !2 into 99 main	SU	0	Completed 1h ago
SU updated test.json Shellunext unext!DA90 request !1 into 99 main	SU	0	Completed 1h ago

https://dev.azure.com/Shellunext1693422200894/IDA_Anushka_Shell/_git/IDA_Anushka_Shell/pullrequests?_a...

Anushka622001.sample-code

IDA_Anushka_Shell +

Overview Boards Repos Pipelines Pipelines Environments Library Test Plans Artifacts Project settings

Triggered by  Anushka622001 View change

Repository and version Tests and coverage

Anushka622001/sample-code Get started

main · 4ebdb85e

Time started and elapsed Related

Just now 0 work items

1m 13s 0 artifacts

Errors 4 Warnings 2

Failed to download Python from the Github Actions python registry (<https://github.com/actions/python-versions>). Error: Error: Could not find Python matching spec 2.7 (x64) in the python-ve...
Job Python27 • Use Python 2.7

Version spec 2.7 for architecture x64 did not match any version in Agent.ToolsDirectory. Versions in /opt/hostedtoolcache: 3.10.13 (x64) 3.11.5 (x64) 3.7.17 (x64) 3.8.18 (x64) 3.9.18 (x64) If this ...
Job Python27 • Use Python 2.7

Failed to download Python from the Github Actions python registry (<https://github.com/actions/python-versions>). Error: Error: Could not find Python matching spec 3.6 (x64) in the python-ve...
Job Python36 • Use Python 3.6

Version spec 3.6 for architecture x64 did not match any version in Agent.ToolsDirectory. Versions in /opt/hostedtoolcache: 3.10.13 (x64) 3.11.5 (x64) 3.7.17 (x64) 3.8.18 (x64) 3.9.18 (x64) If this ...
Job Python36 • Use Python 3.6

Jobs

Name	Status	Duration
Job Python27	Failed	⌚ 4s
Job Python35	Queued	
Job Python36	Failed	⌚ 5s
Job Python37	Success	⌚ 16s

JOURNEY SCRAPBOOK

Custom Bootcamp – Week 5
3 October, 2023 – 6 October, 2023

Sanat Mahajan

Batch 3

DAY 1

1. What is Machine learning and its uses in daily life ?
2. Types of machine learning algorithms :
 - Supervised
 - Unsupervised
 - Semi-supervised
 - Reinforcement
3. Overview of Azure ML Studio
4. Notebooks and Pipelines
5. Azure ML Designer

All workspaces

Home

Model catalog PREVIEW

Authoring

Notebooks

Automated ML

Designer

Prompt flow PREVIEW

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

Compute

Monitoring PREVIEW

Data Labeling

Linked Services

Unext > anushka-ml > Notebooks

Kernel Error Suggestion: Resource Microsoft.MachineLearning.Common.Core.ServiceInvocationException: Service invocation failed! Request: GET exec.nbs.eastus.azuresynapse.net/jupyterApi/versions/1/productTypes/ml/subscriptions...

azureml-getting-star

azureml-getting-start

Edit in VS Code

Compute:

Serverless Spark Compute - Available

Language: PySpark (Python)

Ready (Stop session) Configure session Job details Spark UI

Kernel not found or was deleted. Your state may be lost. Please re-run cells to relaunch.

- Bring data in and prepare it to be used for training
- Train a model for image classification
- Metrics for optimizing your model
- Deploy the model online & test

1. Connect to Azure Machine Learning workspace

Before we dive in the code, you'll need to connect to your workspace. The workspace is the top-level resource for Azure Machine Learning, providing a centralized place to work with all the artifacts you create when you use Azure Machine Learning.

We are using `DefaultAzureCredential` to get access to workspace. `DefaultAzureCredential` should be capable of handling most scenarios. If you want to learn more about other available credentials, go to [set up authentication doc](#), [azure-identity reference doc](#).

Make sure to enter your workspace credentials before you run the script below.

```
1 %pip install azure-ai-ml
2 %pip install azure-identity
[2] ✓ 46 sec - Command executed in 1 ms by Shellunext_169342200894 on 11:04:48 AM, 10/03/23
```

```
Collecting azure-ai-ml
  Downloading azure_ai_ml-1.10.1-py3-none-any.whl (7.0 MB)
    7.0/7.0 MB 104.5 MB/s eta 0:00:00a 0:00:01
```

```
Collecting pydash<6.0.0
  Downloading pydash-5.1.2-py3-none-any.whl (84 kB)
    85.0/85.0 KB 30.9 MB/s eta 0:00:00
```

```
Collecting azure-storage-file-share<13.0.0
  Downloading azure_storage_file_share-12.14.1-py3-none-any.whl (266 kB)
    266.3/266.3 KB 61.0 MB/s eta 0:00:00
```

```
Collecting azure-storage-file-datalake<13.0.0
  Downloading azure_storage_file_datalake-12.13.1-py3-none-any.whl (249 kB)
```

- All workspaces
- Home
- Model catalog PREVIEW
- Authoring
- Notebooks
- Automated ML
- Designer
- Prompt flow PREVIEW
- Assets
- Data
- Jobs
- Components
- Pipelines
- Environments
- Models
- Endpoints
- Manage
- Compute
- Monitoring PREVIEW
- Data Labeling
- Linked Services

Recommendation - Movie Rating Tweets CompletedShareJob overview

Outline

 Type node name, comment or comp...Add filter

Recommendation - Movie Rating T...

Movie Ratings

IMDB Movie Titles

join_data

execute_python_script

remove_duplicate_rows

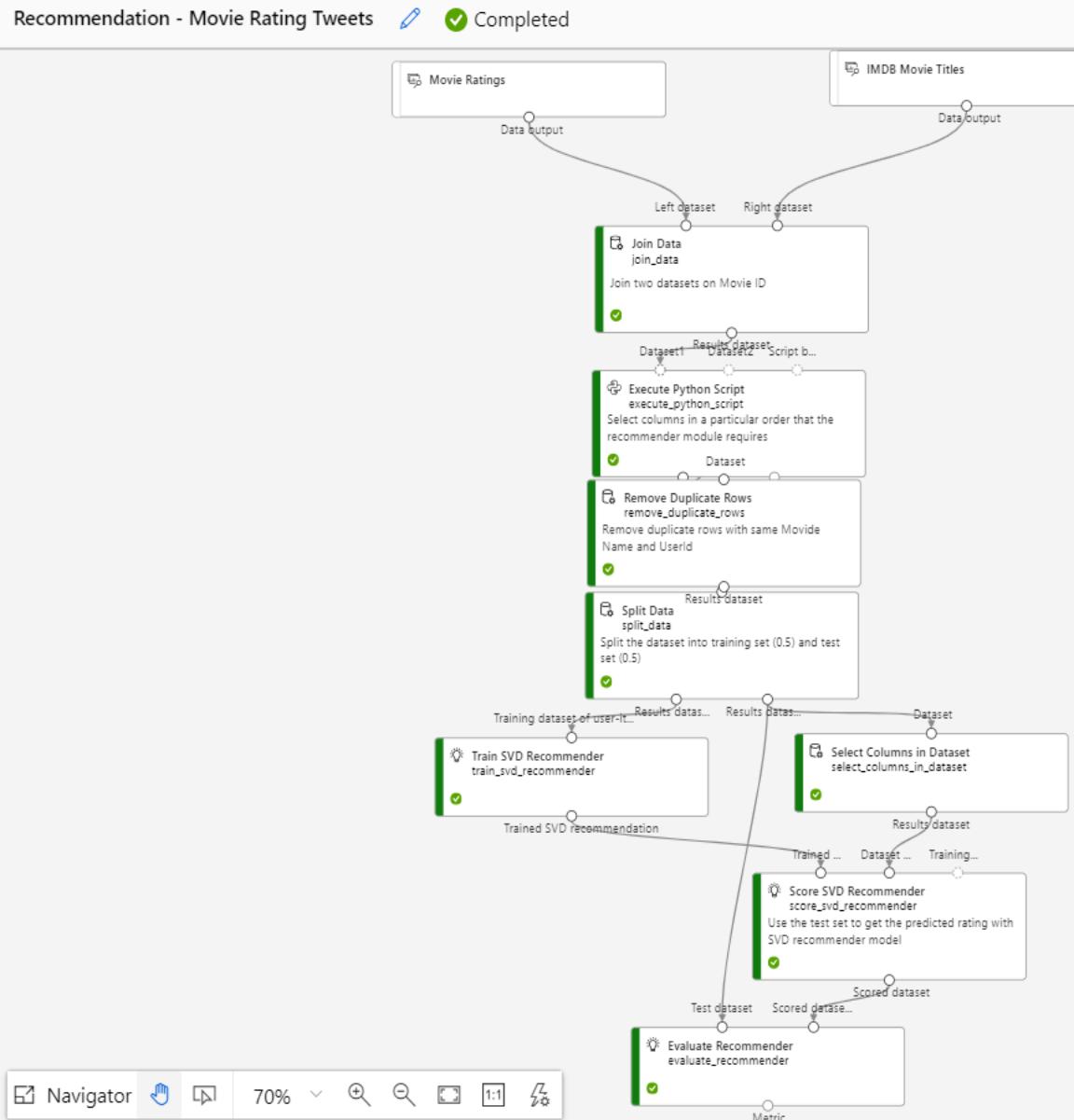
split_data

train_svd_recommender

select_columns_in_dataset

score_svd_recommender

evaluate_recommender



DAY 2

1. Azure ML Studio Pipelines
2. Tune model Hyperparameters
3. Automated ML

Import favorites | For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Azure AI | Machine Learning Studio

Unext > sanatml > Designer > Authoring

Success: Pipeline job has been submitted. View Details

Undo Redo Validate Show lineage Clone AutoSave Configure & Submit

Pipeline-Created-on-10-04-2023

Save Pipeline interface

Dataset

Clean Missing Data clean_missing_data

Cleaned data... Cleaning tran...

Dataset

Boosted Decision Tree Regression boosted_decision_tree_regression

Untrained model

Split Data split_data

Results data... Results data...

Untrain... Training... Optional...

Tune Model Hyperparameters tune_model_hyperparameters

Sweep results Trained best ...

Trained model Dataset

Score Model score_model

Navigator 79% 1:1

All workspaces Home Model catalog (PREVIEW) Notebooks Automated ML Designer Prompt flow (PREVIEW) Assets Data Jobs Components Pipelines Environments Models Endpoints

```
graph TD; A[Clean Missing Data<br>clean_missing_data] --> B[Split Data<br>split_data]; B --> C[Tune Model Hyperparameters<br>tune_model_hyperparameters]; C --> D[Score Model<br>score_model]; D --> E[Trained model]; D --> F[Dataset]
```

Import favorites | For quick access, place your favorites here on the favorites bar. [Manage favorites now](#)

Azure AI | Machine Learning Studio

Unext > sanatml > Jobs > Titanic_PL > frank_flower_88jjstq1

frank_flower_88jjstq1 Running

Overview Data guardrails Models Outputs + logs Child jobs

Refresh Edit and submit (preview) Register model Cancel Delete Compare (preview)

Properties

Status  Running

Model training

Created on Oct 4, 2023 11:35 AM

Start time Oct 4, 2023 11:35 AM

Compute target sanatmlcompute

Name AutoML_93519a4e-7856-4f22-bf4e-bf9e4ac7ce87

Script name --

Inputs

Input name: training_data
Dataset: sdcefgdscv:1

Best model summary

No data

Run summary

Task type Regression [View configuration settings](#)

Featurization Auto

Primary metric Normalized root mean squared error

Experiment name

**THANK
YOU**